

Network-based Pathway Enrichment Analysis with Incomplete Network Information

Jing Ma¹, Ali Shojaie², and George Michailidis³

¹*Department of Biostatistics and Epidemiology, University of Pennsylvania
Perelman School of Medicine, Philadelphia, 19104, USA*

²*Department of Biostatistics, University of Washington, Seattle, 98195, USA*

³*Department of Statistics, University of Florida, Gainesville, 32611, USA*

June 24, 2016

Abstract

Motivation: Pathway enrichment analysis has become a key tool for biomedical researchers to gain insight into the underlying biology of differentially expressed genes, proteins and metabolites. It reduces complexity and provides a system-level view of changes in cellular activity in response to treatments and/or in disease states. Methods that use existing pathway network information have been shown to outperform simpler methods that only take into account pathway membership. However, despite significant progress in understanding the association amongst members of biological pathways, and expansion of data bases containing information about interactions of biomolecules, the existing network information may be incomplete or inaccurate, and is not cell-type or disease condition-specific.

Results: We propose a constrained network estimation framework that combines network estimation based on cell- and condition-specific high-dimensional Omics data with interaction information from existing data bases. The resulting pathway topology information is subsequently used to provide a framework for simultaneous testing of differences in expression levels of pathway members, as well as their interactions. We study the asymptotic properties of the proposed network estimator and the test for pathway enrichment, and investigate its small sample performance in simulated and real data settings.

Availability: The proposed method has been implemented in the R-package `netgsa` available on CRAN.

Contact: jinma@upenn.edu

Supplementary information: Supplementary materials including technical details and dataset are available at *Bioinformatics* online.

1 Introduction

Recent advances in high throughput technologies have transformed biomedical research by enabling comprehensive monitoring of complex biological systems. By profiling the activity of different molecular compartments (genomic, proteomic, metabolomic), one can delineate complex mechanisms that play key roles in biological processes or the development of distinct phenotypes. These technological advances have thus motivated new methodological developments, most notably the adaptation of systems perspectives to analyze biological systems. Pathway analysis represents a key component in the analysis process and has been used successfully in generating new biological hypotheses, as well as in determining whether specific pathways are associated with particular phenotypes. Examples include analysis of pathways involved in initiation and progression of cancer and other complex diseases (Wilson *et al.*, 2010), discovering novel transcriptional effects and co-regulated genes (Green *et al.*, 2011), and understanding the basic biological processes in model organisms (Houstis *et al.*, 2006; Gottwein *et al.*, 2007). See Huang *et al.* (2008) for additional examples of applications.

Pathway analysis methods have evolved since the seminal work by Subramanian *et al.* (2005). As pointed out in the review paper by Khatri *et al.* (2012), earlier techniques such as over-representation analysis (Al-Shahrour *et al.*, 2005), and gene set analysis (Subramanian *et al.*, 2005; Efron and Tibshirani, 2007) treat each pathway as a set of biomolecules. These methods assess whether members of a given pathway have higher than expected levels of activity, either by counting the number of differentially active members, or by also accounting for the relative rankings of pathway members and/or the magnitude of their associations with the phenotype. On the other hand, more recent and statistically powerful methods also account for interactions between biomolecules. These interactions are increasingly available from carefully curated biological databases, such as Kyoto Encyclopedia of Genes and Genomes (KEGG, Kanehisa and Goto, 2000), Reactome (Joshi-Tope *et al.*, 2003), RegulonDB (Huerta *et al.*, 1998) and BioCarta (Nishimura, 2001).

A network topology-based method that exhibits superior statistical power in identifying differential activity of pathways was proposed in Shojaie and Michailidis (2009, 2010). The Network-based Gene Set Analysis (NetGSA) method also allows testing for potential changes in the network structure under different experimental or disease conditions. However, it requires *a priori* knowledge of interactions among pathway members, which, despite rapid progress, remains highly incomplete and occasionally unreliable (see e.g. Zaki *et al.* (2013) and references therein). Moreover, existing network information often determine molecular interactions in the normal state of the cell, and do not provide any insight into condition/disease-specific alterations in interactions amongst components of biological systems.

The increased availability of large sets of high-dimensional Omics data (e.g. from The Cancer Genome Atlas, <http://cancergenome.nih.gov/>), coupled with the development of network estimation techniques based on graphical models (Lauritzen, 1996) offers the possibility to validate and complement existing network information, and to obtain condition-specific estimates of molecular interactions. Such an approach for leveraging existing knowledge to enhance the analysis of low signal-to-noise biological datasets was advocated in Ideker *et al.* (2011).

The first contribution of this paper is the development of a method for constrained network estimation from high-dimensional data, together with establishing the consistency of the resulting estimate. Estimation of high-dimensional networks subject to hard (or soft) constraints on conditional dependence relationships amongst random variables represents a canonical problem in the context of graphical models, and the proposed method for addressing this

problem is of independent interest. By incorporating the condition-specific network estimates from the proposed method into the NetGSA framework, we also provide a rigorous statistical framework for assessing alterations in biological pathways, referred to as *differential network biology* (Ideker and Krogan, 2012).

The proposed framework accounts for two sources of uncertainty: the first concerns the reliability of the external information used for constructing the network estimate from data. The second is the variability of the network estimate, which can impact the pathway enrichment testing procedure. We establish that, under certain regularity conditions, consistent estimates of the network can be obtained, leading, in turn, to an asymptotically most power unbiased test for pathway enrichment analysis. Our theoretical analysis also sheds light into the potential improvements in accuracy and power by directly accounting for the amount of reliable external network information.

A second objective of this study is to scale up the NetGSA estimation algorithm to very large size networks. The main bottleneck in applying the NetGSA methodology arises from the estimation of mixed effects linear parameters—specifically the variance components—for thousands of variables. We develop efficient and stable computational methods for estimation of these parameters based on a profile likelihood approach. In particular, we employ a Cholesky factorization of the covariance matrices to speed up matrix inversions, and use it to develop a stable algorithm based on Newton’s method with backtracking line search (Boyd and Vandenberghe, 2004, page 487) for step size selection. To supply reliable starting points for this algorithm, we further develop an approximate method-of-moment-type estimator.

The proposed methods are illustrated on both metabolomics and gene expression data. For mass spectrometry metabolomics profiling one can obtain good quality measurements for a few hundred metabolites that do not provide complete coverage of the underlying biochemical pathways. The small number of metabolites in each pathway and the incomplete coverage of the metabolites particularly hinder the application of over-representation and gene set analysis methods in this setting. In our experience, only topology-based pathway enrichment analysis methods, such as NetGSA, are capable of reliably delineating pathway activity, as illustrated in Section 4. Further, our investigation of previously analyzed gene expression data set on lung and breast cancer provides new useful insights.

The remainder of the paper is organized as follows. Section 2.1 presents the new method for network estimation under external information constraints and establishes its consistency. Section 2.2 outlines the new computational algorithm for scaling up NetGSA, as well as the inference procedure for both pathway enrichment and differential network analysis. The performance of the developed methodology is evaluated in Section 3 and is examined on real data sets in Section 4.

2 Methods

Gaussian graphical models (Lauritzen, 1996, Chapter 5) are widely used in biological applications to model the interactions among components of biological systems (Dehmer and Emmert-Streib, 2008, chapter 6). Specifically, partial correlation networks are commonly used to model interactions in molecular networks; these networks are represented by an undirected graph $G = (V, E)$ with node set V and edge set E corresponding to biomolecules interactions among them, respectively. The edge set E corresponds to the $p \times p$ precision, or inverse covariance, matrix Ω , whose nonzero elements $\omega_{ii'}$ refer to edges between nodes i and i' , and indicate that i and i' are conditionally dependent given all other nodes in the network. The

magnitude of the partial correlation $\mathbf{A}_{ii'} = -\omega_{ii'} / \sqrt{\omega_{ii}\omega_{i'i'}}$ determines the strength (positive or negative) of the conditional association between the respective nodes. In the sequel, the matrix \mathbf{A} will also be called the weighted adjacency matrix, with $\mathbf{A}_{ii'}$ being the association weight between i and i' .

2.1 Network Estimation Under External Information Constraints

As discussed in Section 1, the availability of large collections of samples for different disease states and biological processes together with carefully curated information of biomolecular interactions enables the estimation of network structures within the setting of Gaussian graphical models. However, the availability of external network information provides a novel and unexplored modification of the corresponding network estimation problem. Denote by E^c the set of node pairs not connected in the network, i.e. $\omega_{ii'} = 0$. Then, the external information can be represented by the following two subsets

$$\begin{aligned} E_1 &= \{(i, i') \in E : i \neq i', \omega_{ii'} \neq 0\}, \\ E_0 &= \{(i, i') \in E^c : i \neq i', \omega_{ii'} = 0\}. \end{aligned}$$

In words, E_1 contains known edges, while E_0 contains node pairs where it is known that no interaction exists between them. Note that $E_1 \subseteq E$ and $E_0 \subseteq E^c$. The external information available in E_1 does not imply exact knowledge of the magnitude of $\omega_{ii'}$ nor $\mathbf{A}_{ii'}$.

Suppose we observe an $m \times p$ data matrix $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_p)$, where each row represents one sample from a p -variate Gaussian distribution $\mathcal{N}(0, \Omega^{-1})$ for a given biological condition (e.g., cancer or normal). Our goal is then to estimate the network structure, or equivalently the precision matrix Ω , subject to external information encoded in E_1 and E_0 . Let $\mathbf{D} = \text{diag}(\Omega)$ represent the diagonal matrix whose diagonal entries are the same as Ω and \mathbf{I}_p be the p -identity matrix. Then, $\mathbf{A} = \mathbf{I}_p - \mathbf{D}^{-1/2}\Omega\mathbf{D}^{-1/2}$ is the partial correlation matrix. When $E_1 = E$ and $E_0 = E^c$, the problem becomes that of covariance selection (Dempster, 1972), which has been studied extensively in the literature. However, to the best of our knowledge, the problem of estimating Ω (and the partial correlation matrix \mathbf{A}) when E_1 and E_0 only contain partial information ($E_1 \subsetneq E$ and $E_0 \subsetneq E^c$) has not been investigated before.

In this section, we assume that the m observations used for estimating condition-specific networks are separate from those used for pathway enrichment analysis (highlighted by the use of \mathbf{Z}_i 's and m to denote the random variables and sample size, respectively). The framework introduced in this section reduces the potential bias in small sample settings, and takes advantage of the additional publicly available samples, in lieu of reliable network information. With large enough samples, network estimation and pathway enrichment can be performed using the same set of samples by incorporating sample splitting strategies. While the problem considered in this section is seemingly similar to matrix completion (Candes and Recht, 2009), the two problems are fundamentally different in nature. In particular, in this setting, matrix completion corresponds to completing the remaining entries from the partially observed $p \times p$ matrix \mathbf{A} , under some structural assumptions on \mathbf{A} , such as low-rankness. On the other hand, in the setting of graphical models, the entries of the weighted adjacency matrix are estimated based on data on the nodes of the graph.

In biological settings, both the structure of the network, as well as strengths of associations may be condition-specific. Therefore, we need to accurately estimate the nonzero entries in Ω to recover both the structure of the network and the strength of associations between nodes. In the absence of any external information, the ℓ_1 -penalized negative log-likelihood estimate of

Ω is obtained by solving

$$\operatorname{argmin}_{\Omega \succ 0} \left\{ \operatorname{trace}(\Omega \hat{\Sigma}) - \log \det \Omega + \lambda \|\Omega\|_1 \right\}, \quad (1)$$

wherein $\hat{\Sigma} = \mathbf{Z}^T \mathbf{Z} / m$ is the empirical covariance matrix of the data, $\|\Omega\|_1 = \sum_{i \neq i'} |\omega_{ii'}|$ denotes the ℓ_1 norm of the parameters, and λ is the regularization parameter. In the presence of external information, the problem can be cast as the following constrained optimization one

$$\min_{\Omega \succ 0} \left\{ \operatorname{trace}(\Omega \hat{\Sigma}) - \log \det \Omega \right\}, \quad (2)$$

subject to $\omega_{ii'} = 0$ for $(i, i') \in E_0$, $\omega_{ii'} \neq 0$ for $(i, i') \in E_1$, and $\sum_{i \neq i', (i, i') \notin E_0 \cup E_1} |\omega_{ii'}| \leq t$.

In the following, we present a two-step procedure to solve the constrained optimization problem (2). The proposed approach combines the neighborhood selection technique (Meinshausen and Bühlmann, 2006) with constrained maximum likelihood estimation. It exploits the fact that the estimated neighbors of each node using neighborhood selection coincide with the nonzero entries of the inverse covariance matrix (Friedman *et al.*, 2008). Specifically, in neighborhood selection the network structure is estimated by finding the optimal set of predictors when regressing the random variable \mathbf{Z}_i corresponding to node $i \in V$ on all other variables, using an l_1 -penalized linear regression. The coefficients for this optimal prediction $\boldsymbol{\theta}^i$ are closely related to the entries of the inverse covariance matrix: for all $i' \neq i$, $\theta_{i'}^i = -\omega_{ii'} / \omega_{ii}$. The set of nonzero coefficients of $\boldsymbol{\theta}^i$ is thus the same as the set of nonzero entries in the row vector of $\omega_{ii'}$ ($i' \neq i$), which defines the set of neighbors of node i .

Let J_1^i and J_0^i denote the subsets $V \setminus i$ for which external information is available: J_1^i is the set of nodes which are known to be in the neighborhood of i , and J_0^i is the set of nodes which are known to be not connected to i . Let \mathbf{Z}_{-i} denote the submatrix obtained by removing the i th column of \mathbf{Z} . Assume all columns of \mathbf{Z} are centered and scaled to have norm 1. Denote by \mathcal{S}_+^p the set of all $p \times p$ positive definite matrices and $\mathcal{S}_E^p = \{\Omega \in \mathbb{R}^{p \times p} : \omega_{ii'} = 0, \text{ for all } (i, i') \notin E \text{ where } i \neq i'\}$. The proposed algorithm proceeds in two steps.

(i) Estimate the network structure \hat{E} . For every node i , find $\hat{\boldsymbol{\theta}}^i$ via the following steps.

- (a) For $i' \in J_0^i$, set $\hat{\theta}_{i'}^i = 0$.
- (b) For $i' \in J_1^i$, find $\hat{\theta}_{i'}^i$ using linear regression

$$\hat{\boldsymbol{\theta}}_{J_1^i}^i = \operatorname{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^{|J_1^i|}} \frac{1}{m} \|\mathbf{Z}_i - \mathbf{Z}_{J_1^i} \boldsymbol{\theta}\|_2^2. \quad (3)$$

- (c) For $i' \in \tilde{J} \equiv V \setminus \{J_1^i \cup J_0^i \cup \{i\}\}$, find $\hat{\theta}_{i'}^i$ using lasso

$$\hat{\boldsymbol{\theta}}_{\tilde{J}}^i = \operatorname{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^{|\tilde{J}|}} \frac{1}{m} \|\mathbf{W}_i - \mathbf{Z}_{\tilde{J}} \boldsymbol{\theta}\|_2^2 + 2\lambda \sum_{i' \in \tilde{J}} |\theta_{i'}^i|, \quad (4)$$

where $\mathbf{W}_i = \mathbf{Z}_i - \mathbf{Z}_{J_1^i} \hat{\boldsymbol{\theta}}_{J_1^i}^i$ is the residual vector after regressing \mathbf{Z}_i on the known connections.

The edge set \hat{E} is estimated to be $\{(i, i') : \hat{\theta}_{i'}^i \neq 0 \text{ OR } \hat{\theta}_i^{i'} \neq 0\}$.

(ii) Given the structure \hat{E} , estimate the inverse covariance matrix $\hat{\Omega}$ by

$$\hat{\Omega} = \operatorname{argmin}_{\Omega \in \mathcal{S}_+^p \cap \mathcal{S}_{\hat{E}}^p} \left\{ \operatorname{trace}(\hat{\Sigma} \Omega) - \log \det \Omega \right\}. \quad (5)$$

Remark 1. In step (i-b) of the algorithm, the coefficients θ^i for known edges have not been penalized in (3). In settings where the external information may be unreliable, we can augment (3) with a lasso penalty $\lambda \sum_{i' \in J_1^i} t_{i'} |\theta_{i'}^i|$, where the penalty weights $t_{i'}$ ($i' \in J_1^i$) allow for different penalization depending on the reliability of existing information.

The second step focuses on estimation of the magnitude of nonzero entries in the precision matrix Ω , given the estimated network topology \hat{E} . The optimization problems in both steps are convex and can be solved efficiently using existing software (e.g., `glmnet` and `glasso` in R).

The proposed estimator enjoys nice theoretical properties under certain regularity conditions. Before presenting the main result, we introduce some additional notations. Let Σ_0 be the true covariance matrix and $\Omega_0 = \Sigma_0^{-1}$. For $i = 1, \dots, p$, denote by $\|\theta^i\|_0 = \#\{i' : \theta_{i'}^i \neq 0\}$ the l_0 norm of θ^i . Write $s = \max_{i=1, \dots, p} \|\theta^i\|_0$ and $S_0 = \sum_{i=1}^p \|\theta^i\|_0$. For a subset $J \subset \{1, \dots, p\}$, let \mathbf{Z}_J be the submatrix obtained by removing the columns whose indices are not in J . We make the following assumptions.

Assumption 1. There exist $\phi_1, \phi_2 > 0$ such that the eigenvalues of Σ_0 are bounded, i.e. $0 < \phi_2 \leq \phi_{\min}(\Sigma_0) \leq \phi_{\max}(\Sigma_0) \leq 1/\phi_1 < \infty$.

Assumption 2. There exists $\kappa(s) > 0$ such that

$$\min_{|J| \leq s} \min_{\substack{\delta \in \mathbb{R}^p \\ \|\delta_{J^c}\|_1 \leq 3\|\delta_J\|_1}} \frac{1}{\sqrt{m}} \frac{\|\mathbf{Z}_J \delta\|_2}{\|\delta_J\|_2} \geq \kappa(s). \quad (6)$$

Assumption 1 is standard in high-dimensional settings. Assumption 2 corresponds to the restricted eigenvalue assumption introduced in Bickel *et al.* (2009), which is presented here for completeness.

Denote by $|E|$ the cardinality of the edge set E . Let $r \equiv (|E_0| + |E_1|)/\{p(p-1)/2\}$ represent the percentage of external network information available. Clearly, $0 \leq r < 1$. We are now ready to state our first result.

Theorem 1. Suppose Assumption 1 holds and Assumption 2 is satisfied with $\kappa(2s)$. For constants $c_1 > 4$ and $0 < k_1 < 1$, assume also that the sample size satisfies

$$m \geq \left\{ \frac{16c_1}{k_1 \phi_1 \kappa^2(2s)} \right\}^2 (1-r) S_0 \log(p-rp), \quad (7)$$

where S_0 is the total number of nonzero parameters excluding the diagonal. Consider $\hat{\Omega}$ defined in (5). Then, with probability at least $1 - 2p^{2-c_1^2/8}$, under appropriately chosen λ , we have

$$\|\hat{\Omega} - \Omega_0\|_2 \leq \|\hat{\Omega} - \Omega_0\|_F = O \left(\sqrt{\frac{S_0 \log(p-rp)}{m}} \right). \quad (8)$$

Remark 2. In addition to the improved sample complexity (7), the convergence rate in (8) indicates an improvement of the order of $\sqrt{S_0 \log(1-r)^{-1}/m}$ in the presence of external information. This improvement is particularly important for our analysis of power properties of NetGSA in Section 2.2.2, which requires norm consistency of adjacency matrix estimation. While consistency can be established using a theoretical analysis similar to graphical lasso (Rothman *et al.*, 2008), our proofs in Section A of the Supplementary Materials utilize the techniques from Bickel *et al.* (2009) and Zhou *et al.* (2011) to characterize the improvement in rates resulting from the external information.

Let \mathbf{A}_0 be the true partial correlation matrix, i.e. $\mathbf{A}_0 = \mathbf{I}_p - \mathbf{D}_0^{-1/2} \Omega_0 \mathbf{D}_0^{-1/2}$, where $\mathbf{D}_0 = \text{diag}(\Omega_0)$. The following corollary is an immediate result of Theorem 1.

Corollary 1. *Let assumptions in Theorem 1 be satisfied. Assume further that $S_0 = o(m/\log(p-rp))$. For $\hat{\Omega}$ defined in (5), let $\hat{\mathbf{A}}$ be the corresponding partial correlation matrix. Then, with probability at least $1 - 2p^{2-c_1^2/8}$, under appropriately chosen λ , we have*

$$\|\hat{\mathbf{A}} - \mathbf{A}_0\|_2 = o(1).$$

Remark 3. *Corollary 1 implies that, under certain regularity conditions, the error in the condition-specific network estimate $\hat{\mathbf{A}}$ is negligible. This proves essential for establishing power properties of NetGSA with estimated network information, as shown in the next section. The proof of Corollary 1 is available in Section A of the Supplementary Materials.*

The tuning parameter λ in the first step of the proposed algorithm is important for selecting the correct structure of the network, which further affects the magnitude of the network interactions in the second step. Accurate estimation of these magnitudes is crucial for topology-based pathway enrichment methods. We propose to select λ using the Bayesian Information Criterion (BIC). Specifically, for a given λ , we define

$$\text{BIC}(\lambda) = \text{trace}(\hat{\Sigma}\hat{\Omega}_\lambda) - \log \det(\hat{\Omega}_\lambda) + \frac{\log(m)}{m} |\hat{E}_\lambda|, \quad (9)$$

where $\hat{\Omega}_\lambda$ is the estimated precision matrix from the data and \hat{E}_λ is the estimated edge set. The optimal tuning parameter is thus $\lambda^* = \arg\min_\lambda \text{BIC}(\lambda)$.

2.2 NetGSA with Estimated Network Information

Next, we discuss how (condition-specific) estimates of bimolecular interactions from Section 2.1 can be incorporated into the NetGSA framework to obtain a rigorous inference procedure for both pathway enrichment and differential network analysis. To this end, we formally define the NetGSA methodology based on undirected Gaussian graphical models in Section 2.2.1. In Section 2.2.2, we discuss how the constrained-network estimation procedure of Section 2.1 can be combined with NetGSA to rigorously infer differential activities of biological pathways, as well as changes in their network structures.

2.2.1 The latent variable model

Consider p genes (proteins/metabolites) whose activity levels across n samples are organized in a $p \times n$ matrix \mathcal{D} . In the framework of NetGSA, the effect of genes (proteins/metabolites) in the network is captured using a latent variable model (Shojaie and Michailidis, 2009, 2010). Denote by \mathbf{Y} an arbitrary column of the data matrix \mathcal{D} . Suppose the observed data can be decomposed into signal, \mathbf{X} , plus noise $\epsilon \sim \mathcal{N}_p(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_p)$, i.e. $\mathbf{Y} = \mathbf{X} + \epsilon$. The latent variable model assumes that the signal \mathbf{X} follows a multivariate normal distribution with partial correlation matrix \mathbf{A} . Based on the connection between linear recursive equations and covariance selection proposed in Wermuth (1980), there exists a lower triangular matrix Λ such that $\Lambda^{-1}\mathbf{X} = \gamma$, where $\gamma \sim \mathcal{N}_p(\boldsymbol{\mu}, \sigma_\gamma^2 \mathbf{I}_p)$ and $\Lambda\Lambda^T = (\mathbf{I}_p - \mathbf{A})^{-1}$. Note that the current version of the NetGSA model differs from the original model in Shojaie and Michailidis (2009, 2010). This difference is primarily manifested through the definition of Λ in the two models:

Λ is defined here based on the *undirected partial correlation network* \mathbf{A} , whereas it was previously defined based on directed (physical) interactions among genes (proteins/metabolites) in Shojaie and Michailidis (2009, 2010).

Assuming that γ and ε are independent, the NetGSA model can then be summarized as

$$\mathbf{Y} = \Lambda\gamma + \varepsilon. \quad (10)$$

The NetGSA methodology allows for more complex models, including time course observations. For expositional clarity, we present the methodology in the setting of two experimental conditions and consider the general case where $\mathbf{A}^{(1)} \neq \mathbf{A}^{(2)}$. Details of NetGSA under multiple conditions can be found in Shojaie and Michailidis (2010) and are applicable for the undirected networks presented in this work. Let $\mathbf{Y}_j^{(k)}$ ($j = 1, \dots, n; k = 1, 2$) be the j -th sample in the expression data under condition k (j th column of data matrix \mathcal{D}), with the first n_1 columns of \mathcal{D} corresponding to condition 1 (control) and the remaining $n_2 = n - n_1$ columns to condition 2 (treatment). Denote by $\Lambda^{(k)}$ the *influence matrix* and $\mu^{(k)}$ the mean vector under condition k . The NetGSA framework considers a latent variable model of the form

$$\begin{aligned} \mathbf{Y}_j^{(1)} &= \Lambda^{(1)}\mu^{(1)} + \Lambda^{(1)}\gamma_j + \varepsilon_j, \quad (j = 1, \dots, n_1), \\ \mathbf{Y}_j^{(2)} &= \Lambda^{(2)}\mu^{(2)} + \Lambda^{(2)}\gamma_j + \varepsilon_j, \quad (j = n_1 + 1, \dots, n). \end{aligned}$$

Here, γ_j is the vector of (unknown) random effects, and ε_j is the vector of random errors. They are independent and normally distributed with mean $\mathbf{0}$ and variances $\sigma_\gamma^2 \mathbf{I}_p$ and $\sigma_\varepsilon^2 \mathbf{I}_p$, respectively.

Inference in NetGSA requires estimation of the mean parameters $\mu^{(1)}$ and $\mu^{(2)}$ and variance components σ_γ^2 and σ_ε^2 . The variance components can be estimated via maximum likelihood or restricted maximum likelihood, which can be computationally demanding for large networks. To extend the applicability of the NetGSA, we consider using Newton's method for estimating the variance parameters based on the profile log-likelihood to improve the computational stability. See Section B of the Supplementary Materials for more details.

2.2.2 Joint pathway enrichment and differential network analysis using NetGSA

To test for enrichment of a pre-specified pathway P , Shojaie and Michailidis (2009) propose the contrast vector (Searle, 1971) $\ell = (-\mathbf{b}\Lambda^{(1)} \odot \mathbf{b}, \mathbf{b}\Lambda^{(2)} \odot \mathbf{b})$, where \mathbf{b} is a row binary vector determining the membership of genes in a pre-specified pathway P and \odot denotes the Hadamard product. The advantage of this contrast vector is that it isolates influences from nodes outside the pathways of interest. Let $\beta = (\mu^{(1)T}, \mu^{(2)T})^T$ be the concatenated vector of means. The null hypothesis of no pathway activity vs the alternative of pathway activation then becomes

$$H_0 : \ell\beta = 0, \quad H_1 : \ell\beta \neq 0. \quad (11)$$

The significance of individual contrast vectors in (11) can be tested using the following Wald test statistic

$$TS = \frac{\ell\hat{\beta}}{\text{SE}(\ell\hat{\beta})}, \quad (12)$$

where $\text{SE}(\ell\hat{\beta})$ represents the standard error of $\ell\hat{\beta}$ and $\hat{\beta}$ is the estimate of β . Both ℓ and $\text{SE}(\ell\hat{\beta})$ depend on the underlying networks, which are estimated using data from the two

experimental conditions. Under the null hypothesis, TS follows approximately a t -distribution whose degrees of freedom can be estimated using the Satterthwaite approximation method (Shojaie and Michailidis, 2010).

The above general framework allows for test of pathway enrichment in arbitrary subnetworks, while automatically adjusting for overlap among pathways. In addition, the above choice of contrast vector ℓ accommodates changes in the network structure. Such changes have been found to play a significant role in development and initiation of complex diseases (Chuang *et al.*, 2012), and NetGSA is currently the only method that systematically combines the changes in expression levels and network structures, when testing for pathway enrichment. However, the applicability of the existing NetGSA framework (Shojaie and Michailidis, 2009, 2010) is limited by the assumption of known network structure (namely $\Lambda^{(k)}$, $k = 1, 2$). In the current framework, we estimate $\Lambda^{(k)}$ ($k = 1, 2$) from data as discussed in Section 2.1. We next show that NetGSA with estimated network information provides valid inference for pathway enrichment and differential network analysis.

For $k = 1, 2$, let $\mathbf{Z}^{(k)}$ of dimension $m_k \times p$ be the data matrix used to separately estimate the partial correlation matrix under condition k . Denote by S_k the number of nonzero off-diagonal entries in the true partial correlation matrix $\mathbf{A}_0^{(k)}$ and by r_k the percentage of available external information. We obtain the following result.

Theorem 2. *Let assumptions in Theorem 1 be satisfied and $S_k = o(m_k / \log(p - r_k p))$ under each condition k ($k = 1, 2$). Consider the inverse covariance matrices $\hat{\Omega}^{(k)}$ estimated from (5) of Section 2.1. Then the test statistic in (12) based on the corresponding networks $\hat{\mathbf{A}}^{(k)}$ is an asymptotically most powerful unbiased test for (11).*

Remark 4. *Theorem 2.1 of Shojaie and Michailidis (2010) says that NetGSA is robust to uncertainty in network information. Specifically, Shojaie and Michailidis (2010) show that if the error in network information $\Delta_{\mathbf{A}_0^{(k)}} = \hat{\mathbf{A}}^{(k)} - \mathbf{A}_0^{(k)}$ satisfies $\|\Delta_{\mathbf{A}_0^{(k)}}\|_2 = o_{\mathbb{P}}(1)$, then NetGSA is an asymptotically most powerful unbiased test for (11). The result in Theorem 2 establishes this property for (partially) estimated networks using the consistency of our proposed network estimation procedure in Theorem 1 and Corollary 1. A detailed proof can be found in Section A of the Supplementary Materials.*

3 Simulation Results

We present two simulation studies to assess the performance of the proposed network estimation procedure, as well as its impact on NetGSA. We refer readers to Section C of the Supplementary Materials for additional simulation scenarios—including validation of Type I errors and settings with a large number of variables p —and subsequent discussion.

Our first experiment is based on an undirected network of size $p = 100$. The network structure is extracted from the DREAM3 challenge (Prill *et al.*, 2010) corresponding to the Ecoli network (labeled Ecoli 1). The pathways of interest are determined through a community detection algorithm based on the leading nonnegative eigenvector of the modularity matrix of the network (Csardi and Nepusz, 2006). Under the null hypothesis, all nodes have the same mean expression values of 1. Under the alternative hypothesis, the mean expression levels of 0%, 30%, 40% and 60% of nodes in subnetworks 1, 3, 5 and 7 are increased by 0.5, respectively.

Our second experiment considers a network of size $p = 160$ with 8 subnetworks of equal sizes, all of which are generated from the same scale-free graph of size 20. To allow for

interactions across subnetworks, there is 20% chance for the hub node in each subnetwork to connect to the hub node in another subnetwork. Mean expression values for all nodes are the same under the null hypothesis. Under the alternative hypothesis, we allow, respectively, 0%, 40%, 60% and 80% of the nodes to have positive mean changes of magnitude 0.5 for subnetworks 1-4. Subnetworks 5-8 follow the same pattern.

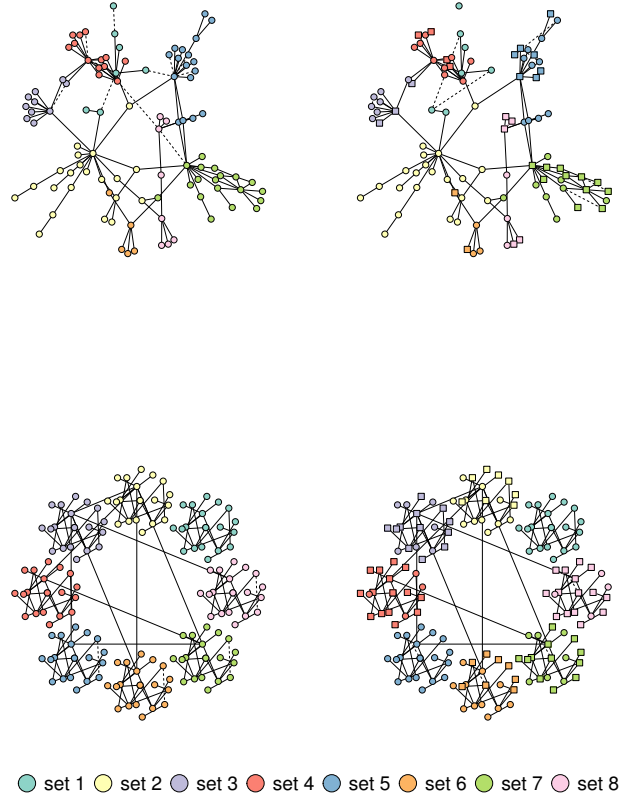


Figure 1: The network and subnetwork topology in experiment 1 under the null (top left) and alternative (top right), and experiment 2 under the null (bottom left) and alternative (bottom right). Dashed lines represent edges that are present in only one condition. Nodes in square are associated with mean changes.

In both experiments, we also allow the structures in four subnetworks under the alternative hypothesis to differ from their null equivalent by a small amount, in order to simultaneously test pathway enrichment and differential networks. Fig. 1 shows the network topologies as well as the structural changes for the chosen subnetworks from the null to the alternative hypothesis in the two experiments. Further, we study the robustness of NetGSA to model misspecification by including scenarios where a proportion (50% for $r = 0.2$ and 20% for $r = 0.8$ in experiment 1, and 60% in experiment 2) of the supplied structural information is incorrectly specified, i.e. they are not present in the true model.

To illustrate how external network information affects the estimation accuracy, we vary the percentage of information r from 0 to 1. When r is less than 1, we estimate the adjacency matrices using the proposed two-step procedure and fill in the nonzero edges with the estimated

weights. When full knowledge of the network topology is given ($r = 1$), we only apply the second step to estimate the edge weights. When there exist misspecified edges in the external information, we use two tuning parameters for network estimation, one for controlling the overall sparsity of the network and the other for correcting the misspecified edges. The optimal tuning parameters are selected over a grid of values using BIC defined in (9).

Table 1 compares the estimated networks with the true model under several deviance measures based on 100 simulation replications; in both experiments, the sample size for both null and alternative hypotheses is $m = 100$. The Matthews correlation coefficients improve significantly as the percentage of external information r increases from 20% to 80%, and the Frobenius norm loss shows a clear decreasing trend, both indicating the improvement in estimation accuracy when more external information is available. In cases where the information is misspecified (denoted by 0.2(m) and 0.8(m)), one can see that the performance of network estimation is not compromised by much after properly selecting the tuning parameters.

Table 1: False positive rate (FPR in percentage), false negative rate (FNR in percentage), Matthews correlation coefficient (MCC) and Frobenius norm loss (Fnorm) for network estimation in experiments 1 and 2

		$p = 100$				$p = 160$			
	r	FPR(%)	FNR(%)	MCC	Fnorm	FPR(%)	FNR(%)	MCC	Fnorm
Null	0.0	9.46	2.78	0.43	0.48	2.94	0.84	0.54	0.36
	0.2	7.64	5.83	0.45	0.46	2.77	1.03	0.55	0.34
	0.8	1.81	1.22	0.75	0.28	1.18	0.02	0.72	0.24
	0.2(m)	7.91	4.85	0.45	0.46	2.76	0.95	0.55	0.34
	0.8(m)	2.29	3.82	0.70	0.31	1.22	0.02	0.71	0.25
Alt	0.0	8.71	1.52	0.44	0.45	2.90	0.88	0.54	0.36
	0.2	7.09	3.82	0.47	0.42	2.73	0.88	0.55	0.35
	0.8	1.80	1.19	0.75	0.25	1.19	1.89	0.71	0.26
	0.2(m)	7.29	2.62	0.47	0.42	2.72	0.78	0.55	0.34
	0.8(m)	2.17	5.50	0.69	0.29	1.22	1.93	0.70	0.27

Next, we examine the performance of NetGSA in detecting pathway enrichment by comparing it with Gene Set Analysis (GSA, Efron and Tibshirani, 2007). GSA tests a competitive null hypothesis and compares the set of genes in the pathway with its complement in terms of association with the phenotype. The underlying model consists of both randomization of the genes and permutation of the samples, which are combined into the idea of ‘restandardization’. This method is later denoted by GSA-c. In addition, we consider GSA with permutation of the samples only, later denoted by GSA-s, since this version of GSA compares the set of genes in the pathway with itself.

Tables 2 and 3 present the estimated powers for each subnetwork in the two experiments from 100 simulation replicates, respectively. Here we use $n_1 = n_2 = 25$ samples for each condition in experiment 1 and $n_1 = n_2 = 40$ in experiment 2, which are different from the datasets used for network estimation. The powers are calculated as the proportion of replicates that show differential changes, based on the false discovery rate (FDR) controlling procedure of Benjamini and Hochberg (1995). To facilitate comparison, different FDR cutoffs are used for GSA and NetGSA to ensure consistent type I error for the first pathway in both experiments. For NetGSA, we look at scenarios when there is 20% and 80% external structural information (with and without misspecification) and use the estimated networks to test enrichment for each

subnetwork. We also include the scenario where the exact networks with correct edge weights are provided, in which case only the variance components and mean expression values are estimated from the mixed linear model. True powers for each subnetwork are calculated by replacing all unknown parameters with their corresponding known values.

Table 2: Powers in experiment 1. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c. Here 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes based on 1000 permutations; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information.

Pathway	$p = 100$							
	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.03	0.03	0.08	0.06	0.15	0.04	0.03	0.02
2	0.08	0.08	0.08	0.06	0.09	0.00	0.09	0.09
3	0.36	0.33	0.43	0.46	0.24	0.00	0.40	0.38
4	0.38	0.26	0.09	0.07	0.26	0.05	0.37	0.24
5	0.91	0.91	0.95	0.97	0.95	0.00	0.92	0.89
6	0.27	0.24	0.24	0.26	0.37	0.00	0.26	0.25
7	0.72	0.80	0.99	0.99	0.98	0.14	0.69	0.86
8	0.45	0.61	0.63	0.57	0.87	0.00	0.51	0.58

For $p = 100$, the results from NetGSA with the exact networks agree with the true powers, indicating low powers for subnetworks 1, 2 and 4, slightly higher powers for 3, 6 and 8, high powers for 5 and 7 due to significant changes in mean expression levels and structures. When the exact networks are unknown, we see clear improvement in the estimated powers for subnetworks 4, 7 and 8 as the percentage of external information increases from 20% to 80%. GSA-s does reasonably well with overestimated powers for subnetwork 8. The last two columns in Table 2 show the estimated powers from NetGSA when the external information is misspecified. For both cases ($r = 0.2$ and $r = 0.8$), the results bear high similarity to those in the first two columns, which suggests that the proposed framework is robust to inaccuracy in network information.

For $p = 160$, the NetGSA estimated powers when 20% external information is available match the true powers reasonably well, with a small underestimation of powers for subnetworks 3, 7 and 8. We note marked improvement in the three corresponding values when the external information increases to 80%. Moreover, NetGSA is able to distinguish subnetworks 5-8 that have both changes in mean values and subnetwork topology from their corresponding counterparts 1-4. When the external information is misspecified, the last two columns indicate that NetGSA still returns valid powers that are comparable to those obtained with correctly specified structural information. GSA-s yields a small overestimation of powers for subnetwork 7.

In both experiments, GSA with randomization of the genes (GSA-c) fails to identify any of the differential subnetworks.

Table 3: Powers in experiment 2. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c. Here 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes based on 1000 permutations; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information.

Pathway	0.2	0.8	E	T	$p = 160$		0.2(m)	0.8(m)
					GSA-s	GSA-c		
1	0.04	0.06	0.02	0.05	0.06	0.02	0.04	0.06
2	0.37	0.36	0.30	0.36	0.41	0.00	0.36	0.36
3	0.88	0.94	0.96	0.99	0.99	0.00	0.89	0.94
4	0.97	0.99	1.00	1.00	1.00	0.23	0.97	0.99
5	0.36	0.25	0.11	0.11	0.13	0.15	0.35	0.25
6	0.38	0.27	0.03	0.07	0.26	0.01	0.35	0.27
7	0.66	0.72	0.92	0.92	1.00	0.00	0.67	0.72
8	0.90	0.95	1.00	1.00	1.00	0.13	0.91	0.94

4 Applications to Metabolomics and Genomics Data

We apply NetGSA to three Omics data sets to demonstrate its potential in revealing biological insights. In all three studies, the p -values were corrected for multiple comparisons using the false discovery rate control procedure proposed in Benjamini and Yekutieli (2001) to account for the dependency among KEGG pathways.

Our first application is based on the metabolomics data set from (Putluri *et al.*, 2011) to examine changes in metabolic profiles associated with bladder cancer using untargeted mass spectrometry data acquisition strategy. The data consists of 31 cancer and 27 benign tissue samples and 63 detected metabolites. Here we focused on estimating the network of metabolic interactions, enhanced by information gleaned from KEGG (Kanehisa and Goto, 2000). For each condition, we used the BIC criterion to select the tuning parameter λ . At the optimal λ , we applied the proposed network estimation procedure to identify the metabolic network; see Fig. 2 for an illustration of the estimated networks for the cancer and benign classes, respectively. It can be seen that there are numerous interactions between pathways that describe energy metabolism in the cancer state, due to the greater need of cancer cells for energy.

We tested for differential activity of biochemical pathways extracted from KEGG using the same set of data. Shown in Table 4 are estimated p -values after false discovery rate correction with a q -value of 0.01 for the significant pathways selected from NetGSA. These identified pathways include those that describe altered utilization of amino acids and their aromatic counterparts, as well as metabolism of fatty acids and intermediates of tricarboxylic acid cycle (TCA) which were followed up for biological insights in the original study by Putluri *et al.* (2011). Among the selected pathways, fatty acid biosynthesis is not identified by GSA-s. Interestingly, GSA-c fails to report any pathway as being significantly enriched. This again confirms our hypothesis that incorporating pathway topology information allows sophisticated enrichment methods in detecting important regulatory pathways.

The second data set (Subramanian *et al.*, 2005) consists of gene expression profiles of 5217 genes for 62 normal and 24 lung cancer patients. We considered 47 KEGG pathways of size at least 5 that describe signaling and biochemical mechanisms and excluded genes that either

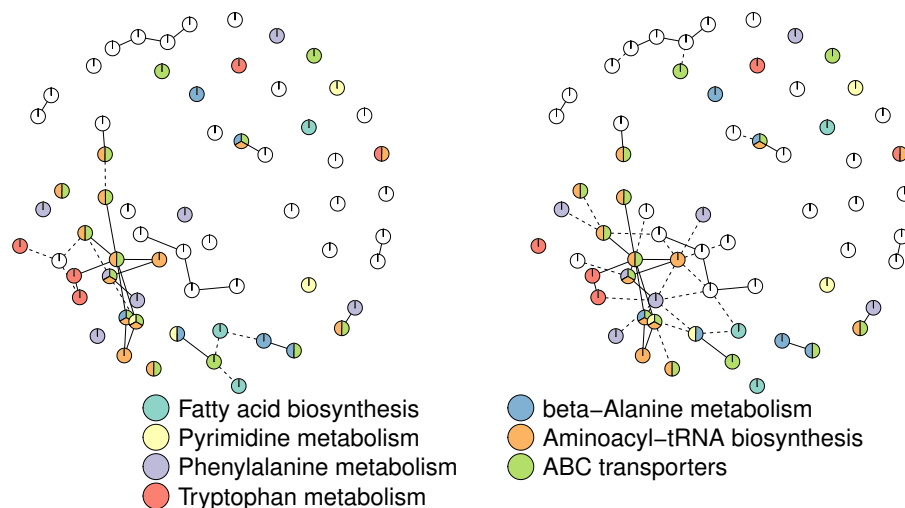


Figure 2: The estimated network topology and enriched pathways in the metabolomics study for the benign class (left) and cancer class (right). Dashed lines represent edges that are present in only one class. Nodes in multiple colors are present in multiple pathways.

are not present in the 47 pathways, or without recorded network information. The number of genes that remain for pathway enrichment analysis is 303. Based on the external topology information from the BioGRID database, we applied the proposed network estimation procedure coupled with BIC to estimate the underlying interaction networks for both normal and cancer conditions. We then explored whether GSA and NetGSA with the estimated networks are able to detect enriched pathways using the same data set. After correcting for multiple comparisons, using a false discovery rate of $q^* = 0.01$, none of the three methods identifies any pathway as being significantly differentially enriched. The lack of statistical power in obtaining differential pathways was also noted in the original paper of (Subramanian *et al.*, 2005); see Table 9 of the Supplementary Materials for the complete list of FDR adjusted p -values. Nevertheless, using NetGSA and a relaxed FDR cutoff threshold of 0.30 (similar to the strategy adopted in (Subramanian *et al.*, 2005)), we obtain the following top three ranked signaling pathways: Jak-STAT, p53 and Wnt. All three are implicated in lung cancer, although the latter two are also implicated in multiple other types of human malignancies. However, the Jak/STAT pathway has been recently shown to play a key role in non small cell lung cancer cells (Song *et al.*, 2011).

Our third and final application is based on a data set from The Cancer Genome Atlas (TCGA, 2012). The data set contains RNA-seq measurements for 17296 genes from 1033 breast cancer specimens, including ER positive, ER negative and other unevaluated cases. As in the previous gene microarray study, the external network information is extracted from the BioGRID database. We focused on a subset of the genes that have recorded network information and are present in KEGG pathways with at least 5 members. This leaves for further consideration 800 genes with 403 samples from the ER positive and 117 from the ER

Table 4: p -values for the pathways in the metabolomics study, with false discovery rate correction at $q^* = 0.01$. Here 0.00 represents a zero p -value produced out of finite permutations. GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes based on 3000 permutations.

Pathway	NetGSA	GSA-s	GSA-c
Tryptophan metabolism	$3e^{-5}$	0.00	1.00
beta-Alanine metabolism	$3e^{-5}$	0.00	1.00
Aminoacyl-tRNA biosynthesis	$2e^{-4}$	0.00	1.00
ABC transporters	$4e^{-4}$	0.00	1.00
Fatty acid biosynthesis	$2e^{-3}$	1.00	1.00
Pyrimidine metabolism	$2e^{-3}$	0.00	1.00
Phenylalanine metabolism	$4e^{-3}$	0.00	1.00

negative classes, spanning over 45 KEGG pathways. We then applied the constrained network estimation procedure with the tuning parameter selected via BIC in (9) to obtain the partial correlation networks for the ER positive and ER negative classes, respectively. Due to the large number of variables, visualization of the estimated networks at the individual gene level is challenging. Instead, we examine the interactions among pathways in Fig. 3 to gain insight into their co-regulation behavior. The weighted pathway level network is defined as follows. Let each node in the network represent one pathway, with size proportional to the size of the corresponding pathway. A weighted edge between two pathways P_1 and P_2 is defined as the number of nonzero partial correlations between genes in P_1 and those in P_2 (normalized by the sizes of the two pathways). Links visualized in Fig. 3 are the top 5% of the weighted edges, where ranking is based on edge weights.

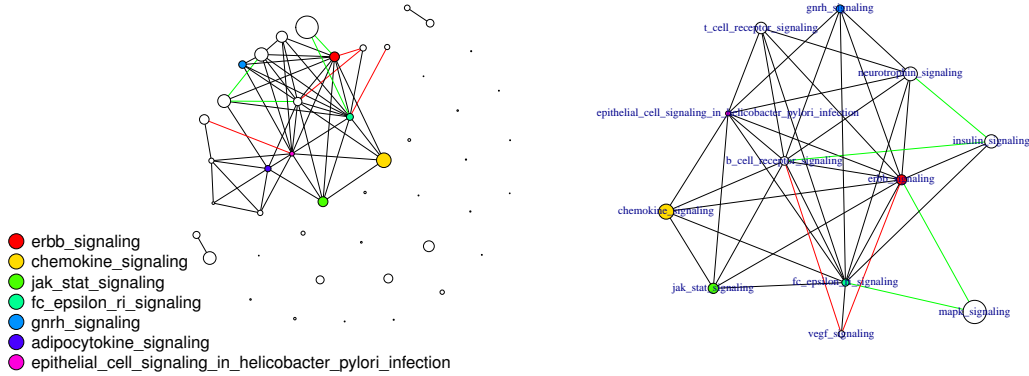


Figure 3: The estimated pathway topology (left) and the subnetwork surrounding the ErbB pathway (right) in the TCGA cancer study. Edges in black are present in both classes, whereas red and green edges are only present in ER positive and ER negative class, respectively. Node size is proportional to the size (number of genes) of the corresponding pathway.

Table 5 presents the false discovery rate corrected p -values for the selected differential pathways using NetGSA based on the estimated partial correlation networks, as well as GSA-c and GSA-s. The complete table is presented in Section D of the Supplementary Materials. At $q^* = 0.01$, NetGSA reports 25 out of the 45 KEGG pathways as significantly enriched, whereas GSA either rejects the null for all pathways (GSA-c) or fails to reject any pathway (GSA-s). Selected differential pathways identified by NetGSA are also highlighted in Fig. 3. Of particular interest is the set of connected, enriched pathways centered around the ErbB pathway in the right panel of Fig. 3. This pathway contains receptors that signal through various pathways to regulate cell proliferation, migration, differentiation, apoptosis, and cell motility and play a key role in breast cancer (Howe and Brown, 2011), although its role in breast carcinogenesis not very well understood. Note that the Jak-STAT pathway is downstream of the ErbB one and can be activated by key epidermal growth factor receptors in the former to create signaling cascades (Henson and Gibson, 2006). Further, the GnRH signaling pathway has been reported to interact with the ErbB pathway receptors (Morgan *et al.*, 2011). All these interconnected pathways are related to receptors that have been implicated in various studies with over-expression in the ER negative class and hence faster tumor growth and poorer clinical outcomes.

5 Discussion

This paper introduces a constrained partial correlation network estimation method that seamlessly incorporates externally available interaction information for genes and other biomolecules. The end product is a reliable condition-specific estimate of the underlying networks. The resulting estimated network structures are then used for network-based pathway enrichment analysis. For the purpose of constrained network estimation, one might also try the one-step constrained maximum likelihood estimation (a functionality offered in the R-package `glasso`) to recover the underlying partial correlation network. However, this one-step approach requires sophisticated specification of the tuning parameters at positions for which structural information is available, and can be challenging to implement in practice.

Two sources of uncertainty can be identified in the proposed framework: one from the reliability of the external database information in the network estimation procedure and the other from the uncertainty regarding the estimated network itself, as well as how it propagates into the NetGSA testing procedure. As discussed in Remark 1, the proposed method can conveniently accommodate the first source of uncertainty by incorporating a non-zero penalty on parameters that are uncertain. Further, as shown in Theorem 2, the proposed test via the extended NetGSA framework is asymptotically unbiased and most powerful, given the consistency of the estimated network, and hence accounts for the second source of uncertainty. Nevertheless, in finite samples as the numerical work in Section C of the Supplementary Materials illustrates, Type I errors may be slightly off in the presence of numerous errors in the estimated network (either due to misspecification of the external information or lack of samples for accurate estimation). The topic of dealing with network estimation errors and possible ways to address it is discussed in Narayan and Allen (2016).

Finally, the current framework of NetGSA uses the Cholesky decomposition of the covariance matrix of the underlying network. It is natural to ask whether the order of the variables affects the result of enrichment analysis. In simulations and the real data analyses, we find that the estimated powers/ p -values from NetGSA are comparable after permutation of the variables.

Table 5: p -values for the differential pathways in the TCGA data, with false discovery rate correction at $q^* = 0.01$. Here 0.00 represents a zero p -value produced out of finite permutations. GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes based on 3000 permutations.

Pathway	NetGSA	GSA-s	GSA-c
Epithelial cell signaling in Helicobacter pylori infection	$5e^{-95}$	0.00	1.00
Cell cycle	$2e^{-47}$	0.00	1.00
Galactose metabolism	$3e^{-31}$	0.00	1.00
Glutathione metabolism	$1e^{-27}$	0.00	1.00
NOD-like receptor signaling pathway	$1e^{-24}$	0.00	1.00
Pyrimidine metabolism	$4e^{-23}$	0.00	1.00
Cysteine and methionine metabolism	$1e^{-22}$	0.00	1.00
Starch and sucrose metabolism	$1e^{-18}$	0.00	1.00
Toll-like receptor signaling pathway	$1e^{-18}$	0.00	1.00
Glycolysis / Gluconeogenesis	$3e^{-17}$	0.00	1.00
Jak-STAT signaling pathway	$9e^{-15}$	0.00	1.00
Chemokine signaling pathway	$3e^{-14}$	0.00	1.00
ErbB signaling pathway	$7e^{-13}$	0.00	1.00
p53 signaling pathway	$7e^{-12}$	0.00	1.00
Hedgehog signaling pathway	$5e^{-10}$	0.00	1.00
beta-Alanine metabolism	$1e^{-7}$	0.00	1.00
Fc epsilon RI signaling pathway	$5e^{-7}$	0.00	1.00
Fructose and mannose metabolism	$2e^{-6}$	0.00	1.00
Pentose phosphate pathway	$2e^{-6}$	0.00	1.00
PPAR signaling pathway	$5e^{-6}$	0.00	1.00
Adipocytokine signaling pathway	$4e^{-5}$	0.00	1.00
Purine metabolism	$6e^{-5}$	0.00	1.00
Valine, leucine and isoleucine degradation	$5e^{-4}$	$1e^{-3}$	1.00
GnRH signaling pathway	$2e^{-3}$	0.00	1.00
TGF-beta signaling pathway	$3e^{-3}$	0.00	1.00

Acknowledgements

The authors would like to thank the reviewers for many constructive comments and suggestions.

References

- Al-Shahrour, F., Díaz-Uriarte, R., and Dopazo, J. (2005). Discovering molecular functions significantly related to phenotypes by combining gene expression data and biological information. *Bioinformatics*, **21**(13), 2988–2993.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(1), 289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

- Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Annals of Statistics*, **36**(1), 199–227.
- Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, **37**(4), 1705–1732.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, **16**(5), 1190–1208.
- Candes, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, **9**, 717–772.
- Chuang, H.-Y., Rassenti, L., Salcedo, M., Licon, K., Kohlmann, A., Haferlach, T., Foà, R., Ideker, T., and Kipps, T. J. (2012). Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood*, **120**(13), 2639–2649.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.
- Dehmer, M. and Emmert-Streib, F. (2008). *Analysis of Microarray Data: a Network-Based Approach*. John Wiley & Sons.
- Dempster, A. P. (1972). Covariance selection. *Biometrics*, **28**(1), 157–175.
- Efron, B. and Tibshirani, R. (2007). On testing the significance of sets of genes. *Annals of Applied Statistics*, **1**(1), 107–129.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**(3), 432–441.
- Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W. H., Chi, J.-T. A., Braich, R., Manoharan, M., Soutschek, J., Ohler, U., *et al.* (2007). A viral microrna functions as an orthologue of cellular mir-155. *Nature*, **450**(7172), 1096–1099.
- Green, M. R., Monti, S., Dalla-Favera, R., Pasqualucci, L., Walsh, N. C., Schmidt-Supprian, M., Kutok, J. L., Rodig, S. J., Neuberg, D. S., Rajewsky, K., *et al.* (2011). Signatures of murine b-cell development implicate *yy1* as a regulator of the germinal center-specific program. *Proceedings of the National Academy of Sciences*, **108**(7), 2873–2878.
- Haberman, S. J. (1989). Concavity and estimation. *Annals of Statistics*, **17**(4), 1631–1661.
- Henson, E. S. and Gibson, S. B. (2006). Surviving cell death through epidermal growth factor (egf) signal transduction pathways: implications for cancer therapy. *Cellular signalling*, **18**(12), 2089–2097.
- Houstis, N., Rosen, E. D., and Lander, E. S. (2006). Reactive oxygen species have a causal role in multiple forms of insulin resistance. *Nature*, **440**(7086), 944–948.
- Howe, L. R. and Brown, P. H. (2011). Targeting the her/egfr/erbb family to prevent breast cancer. *Cancer prevention research*, **4**(8), 1149–1157.

- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols*, **4**(1), 44–57.
- Huerta, A. M., Salgado, H., Thieffry, D., and Collado-Vides, J. (1998). Regulondb: A database on transcriptional regulation in escherichia coli. *Nucleic Acids Research*, **26**(1), 55–59.
- Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Molecular Systems Biology*, **8**(1), 565.
- Ideker, T., Dutkowski, J., and Hood, L. (2011). Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell*, **144**(6), 860–863.
- Joshi-Tope, G., Vastrik, I., Gopinath, G., Matthews, L., Schmidt, E., Gillespie, M., D’Eustachio, P., Jassal, B., Lewis, S., Wu, G., Birney, E., and Stein, L. (2003). The genome knowledgebase: A resource for biologists and bioinformaticists. *Cold Spring Harbor Symposium on Quantitative Biology*, **68**, 237–244.
- Kanehisa, M. and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**(1), 27–30.
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*, **8**(2), e1002375.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
- Lindstrom, M. J. and Bates, D. M. (1988). Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83**(404), 1014–1022.
- Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *Annals of Statistics*, **34**(3), 1436–1462.
- Mirsky, L. (1975). A trace inequality of john von neumann. *Monatshefte für Mathematik*, **79**(4), 303–306.
- Morgan, K., Meyer, C., Miller, N., Sims, A. H., Cagnan, I., Faratian, D., Harrison, D. J., Millar, R. P., and Langdon, S. P. (2011). Gnrh receptor activation competes at a low level with growth signaling in stably transfected human breast cell lines. *BMC cancer*, **11**(1), 476.
- Narayan, M. and Allen, G. I. (2016). Mixed effects models to find differences in multi-subject functional connectivity. *Frontiers in Neuroscience*, **10**(108).
- Nishimura, D. (2001). Biocarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient*, **2**(3), 117–120.
- Prill, R. J., Marbach, D., Saez-Rodriguez, J., Sorger, P. K., Alexopoulos, L. G., Xue, X., Clarke, N. D., Altan-Bonnet, G., and Stolovitzky, G. (2010). Towards a rigorous assessment of systems biology models: the dream3 challenges. *PloS one*, **5**(2), e9202.

- Putluri, N., Shojaie, A., Vasu, V. T., Vareed, S. K., Nalluri, S., Putluri, V., Thangjam, G. S., Panzitt, K., Tallman, C. T., Butler, C., *et al.* (2011). Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer Research*, **71**(24), 7376–7386.
- Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, **2**, 494–515.
- Searle, S. (1971). *Linear Models*. John Wiley & Sons.
- Shojaie, A. and Michailidis, G. (2009). Analysis of gene sets based on the underlying regulatory network. *Journal of Computational Biology*, **16**(3), 407–426.
- Shojaie, A. and Michailidis, G. (2010). Network enrichment analysis in complex experiments. *Statistical Applications in Genetics and Molecular Biology*, **9**(1), Article 22.
- Song, L., Rawal, B., Nemeth, J. A., and Haura, E. B. (2011). Jak1 activates stat3 activity in non-small-cell lung cancer cells and il-6 neutralizing antibodies can suppress jak1-stat3 signaling. *Molecular cancer therapeutics*, **10**(3), 481–494.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43), 15545–15550.
- TCGA (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418), 61–70.
- Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *Journal of the American Statistical Association*, **75**(372), 963–972.
- Wilson, B. G., Wang, X., Shen, X., McKenna, E. S., Lemieux, M. E., Cho, Y.-J., Koellhoffer, E. C., Pomeroy, S. L., Orkin, S. H., and Roberts, C. W. (2010). Epigenetic antagonism between polycomb and swi/snf complexes during oncogenic transformation. *Cancer cell*, **18**(4), 316–328.
- Zaki, N., Efimov, D., and Berengueres, J. (2013). Protein complex detection using interaction reliability assessment and weighted clustering coefficient. *BMC Bioinformatics*, **14**(1), 163.
- Zhou, S., Rutimann, P., Xu, M., and Bühlmann, P. (2011). High-dimensional covariance estimation based on gaussian graphical models. *Journal of Machine Learning Research*, **12**, 2975–3026.

Supplementary materials to “Network-based pathway enrichment analysis with incomplete network information”

A Theoretical Analysis and Proofs

We first introduce additional notation needed in the remainder. Define $\tilde{\Omega}_0 = \text{diag}(\Omega_0) + \Omega_{0, E \cap \hat{E}}$, where E and \hat{E} are the true and the estimated edge set, respectively. By definition, $\tilde{\Omega}_0$ and Ω_0 will be different at position (i, i') only when the edge (i, i') is falsely rejected. In the following, we first derive an upper bound for the size of \hat{E} and $\|\tilde{\Omega}_0 - \Omega_0\|_F$. For the ease of presentation, we drop the superscript i for sets J_0 and J_1 in the i th regression, but they should be understood as J_0^i and J_1^i , respectively.

The following lemma is needed in the proof of Theorem A.1 below.

Lemma A.1. *For $i = 1, \dots, p$, denote by $\xi^i = \mathbf{Z}_i - \sum_{i' \neq i} \theta_{i'}^i \mathbf{Z}_{i'}$, where θ^i is the optimal prediction coefficient vector in the i th regression. Consider the event*

$$\mathcal{F}_i := \left\{ \mathbf{Z} : \frac{1}{m} \|\mathbf{Z}_{-i}^T \xi^i\|_\infty \leq \frac{c_1}{2} \sqrt{\frac{\log(p - rp)}{m\omega_{0,ii}}} \right\}$$

with a constant $c_1 > 4$, where $\omega_{0,ii}$ is the i th diagonal element of the true inverse covariance matrix Ω_0 . Define the event $\mathcal{F} = \bigcap_{i=1}^p \mathcal{F}_i$. Then $\mathbb{P}(\mathcal{F}) \geq 1 - 2p^{2-c_1^2/8}$.

The proof of Lemma A.1 will be provided shortly. Denote by Λ_{\max} the maximal eigenvalue of $\mathbf{Z}^T \mathbf{Z} / m$. Conditioning on the event \mathcal{F} , we have the following results on controlling the size of \hat{E} and the Frobenius norm of the deviance, $\|\tilde{\Omega}_0 - \Omega_0\|_F$.

Theorem A.1. *Suppose the conditions in Theorem 2.2 are satisfied. Then on event \mathcal{F} , for appropriately chosen λ , we have*

$$|\hat{E}| \leq \frac{64\Lambda_{\max}}{\kappa^2(s)}(1-r)S_0 + rS_0, \quad (\text{A.1})$$

and

$$\|\tilde{\Omega}_0 - \Omega_0\|_F \leq c_3 \sqrt{\frac{S_0 \log(p - rp)}{m}} \leq k_1 \phi_1, \quad (\text{A.2})$$

where $c_3 = 16c_1 \sqrt{(1-r)/\kappa^2(2s)}$.

Remark 5. *The result indicates that the cardinality of the estimated edge set is upper bounded by a function of r , the percentage of the external information. The bound for $|\hat{E}|$ also depends on the restricted eigenvalue $\kappa(s)$, which is necessarily positive by the assumption that $\kappa(2s) > 0$. Two extreme cases occur when (i) $r = 0$, i.e. we do not observe any external information, thus reducing problem (2.4) to the original neighborhood selection in Meinshausen and Bühlmann (2006); (ii) $r = 1$, i.e. the exact network topology is known and hence $\hat{E} = E$. On the other hand, the upper bound for $\|\tilde{\Omega}_0 - \Omega_0\|_F$ decreases as r increases, i.e. when more external information becomes available. However, since the coefficients also need to be estimated, this deviance always stays positive, even when $r = 1$.*

Proof of Theorem A.1. Recall $\tilde{J} = V \setminus \{J_1 \cup J_0 \cup \{i\}\}$ is the set of indices for which there is no information available. Denote by $\mathbf{P}_{J_1} = \mathbf{Z}_{J_1}(\mathbf{Z}_{J_1}^T \mathbf{Z}_{J_1})^{-1} \mathbf{Z}_{J_1}^T$ the projection onto the column space of \mathbf{Z}_{J_1} . It is easy to see that the problem (2.4) is equivalent to solving

$$\min_{\boldsymbol{\theta}_{\tilde{J}}} \frac{1}{m} \|(\mathbf{I}_p - \mathbf{P}_{J_1})\mathbf{Z}_i - \mathbf{Z}_{\tilde{J}}\boldsymbol{\theta}_{\tilde{J}}\|_2^2 + 2\lambda \|\boldsymbol{\theta}_{\tilde{J}}\|_1. \quad (\text{A.3})$$

To bound \hat{E} and $\|\tilde{\Omega}_0 - \Omega_0\|_F$, it suffices to focus mainly on the set \tilde{J} , as false positive and negative errors will only occur on this set.

Denote by s_1^i and s^i , respectively, the number of known ones and the number of nonzero coordinates after excluding the known ones in the i th regression, and $s = \max_{i=1, \dots, p} (s_1^i + s^i)$. If \mathbf{Z} satisfies the restricted eigenvalue condition in Assumption 2 with $\kappa(2s) > 0$, then $\mathbf{Z}_{\tilde{J}}$ satisfies the same assumption with $\kappa(2s^i) \geq \kappa(2s) > 0$ for $s^i \leq s$. Moreover, $\kappa(s^i) \geq \kappa(s) \geq \kappa(2s) > 0$. Let $\hat{\boldsymbol{\theta}}_{\tilde{J}}^i$ be the lasso estimator in (A.3) with

$$\lambda = c_1 \sqrt{\frac{\log(p - rp)}{m\omega_{0,ii}}} \quad (\text{A.4})$$

for $c_1 > 4$. Conditioning on the event \mathcal{F} , we can invoke Theorem 7.2 of Bickel *et al.* (2009) and obtain simultaneously for all i ,

$$\|\hat{\boldsymbol{\theta}}_{\tilde{J}}^i\|_0 \leq \frac{64\Lambda_{\max}}{\kappa^2(s^i)} s^i, \quad (\text{A.5})$$

and

$$\|\hat{\boldsymbol{\theta}}_{\tilde{J}}^i - \boldsymbol{\theta}_{\tilde{J}}^i\|_2 \leq \frac{16c_1}{\omega_{0,ii}\kappa^2(2s^i)} \sqrt{\frac{s^i \log(p - rp)}{m}}. \quad (\text{A.6})$$

Combining (A.5) with the number of known edges s_1^i as given in J_1^i , we get

$$|\hat{E}| \leq \sum_{i=1}^p \{ \|\hat{\boldsymbol{\theta}}_{\tilde{J}}^i\|_0 + |J_1^i| \} \leq \frac{64\Lambda_{\max}}{\kappa^2(s)} \sum_{i=1}^p s^i + \sum_{i=1}^p s_1^i.$$

The upper bound in (A.1) follows immediately, since by definition the number of known and unknown edges are $\sum_{i=1}^p s_1^i = rS_0$ and $\sum_{i=1}^p s^i = (1 - r)S_0$, respectively.

To bound $\|\tilde{\Omega}_0 - \Omega_0\|_F$, recall that for every $i' \neq i$, $\omega_{0,ii'} = -\theta_{i'}^i \omega_{0,ii}$. Using the bound in (A.6), we have

$$\begin{aligned} \|\tilde{\Omega}_0 - \Omega_0\|_F^2 &= \sum_{i=1}^p \sum_{i' \in J(\theta^i) \cap J(\hat{\theta}^i)^c} (\theta_{i'}^i \omega_{0,ii})^2 = \sum_{i=1}^p \omega_{0,ii}^2 \sum_{i' \in J(\theta^i) \cap J(\hat{\theta}^i)^c} |\theta_{i'}^i - \hat{\theta}_{i'}^i|^2 \\ &\leq \sum_{i=1}^p \omega_{0,ii}^2 \|\boldsymbol{\theta}_{\tilde{J}}^i - \hat{\boldsymbol{\theta}}_{\tilde{J}}^i\|_2^2 \leq \left\{ \frac{16c_1}{\kappa^2(2s)} \right\}^2 \frac{(1 - r)S_0 \log(p - rp)}{m}. \end{aligned}$$

The last inequality in (A.2) follows from condition (2.7) in Theorem 2.2. \square

Proof of Lemma A.1. For every i , it is easy to verify that $\boldsymbol{\xi}^i$ is normally distributed with mean $\mathbf{0}$ and variance $1/\omega_{0,ii} \mathbf{I}_m$. Define random variables $\Upsilon_{ii'} = (\omega_{0,ii}/m)^{1/2} \mathbf{Z}_{i'}^T \boldsymbol{\xi}^i$ for $i' \neq i$. Then,

$\mathbf{Z}_{i'}^T \mathbf{Z}_{i'}/m = 1$ implies that $\Upsilon_{ii'} \sim \mathcal{N}(0, 1)$. Let λ be defined as in (A.4). Using an elementary bound on the tails of Gaussian distributions,

$$\begin{aligned} \mathbb{P}(\mathcal{F}^c) &\leq \sum_{i=1}^p \sum_{i' \neq i} \mathbb{P}(\{|\mathbf{Z}_{i'}^T \boldsymbol{\xi}^i|/m > \lambda/2\}) \\ &\leq \sum_{i=1}^p \sum_{i' \neq i} \mathbb{P}\left(|\Upsilon_{ii'}| > (m\omega_{0,ii})^{1/2} \lambda/2\right) \leq \sum_{i=1}^p \sum_{i' \neq i} 2 \exp\{-m\omega_{0,ii} \lambda^2/8\} \\ &\leq 2p(p-1) \exp\{-c_1^2 \log(p-rp)/8\} \leq 2p^{2-c_1^2/8}. \end{aligned}$$

Therefore, $\mathbb{P}(\mathcal{F}) \geq 1 - 2p^{2-c_1^2/8}$. \square

With Lemma A.1 and Theorem A.1, we are ready to prove our main result in Theorem 2.2. The following proof is adapted from Zhou *et al.* (2011).

Proof of Theorem 2.2. Consider $\hat{\Omega}$ defined in (2.5). It suffices to show that on the event \mathcal{F}

$$\|\hat{\Omega} - \tilde{\Omega}_0\|_F = O\left(\sqrt{\frac{S_0 \log(p-rp)}{m}}\right),$$

since by triangle inequality and Theorem A.1, we can conclude

$$\|\hat{\Omega} - \Omega_0\|_F \leq \|\hat{\Omega} - \tilde{\Omega}_0\|_F + \|\tilde{\Omega}_0 - \Omega_0\|_F \leq O\left(\sqrt{\frac{S_0 \log(p-rp)}{m}}\right).$$

Denote $\tilde{\Sigma}_0 = \tilde{\Omega}_0^{-1}$, which is positive definite since by Theorem A.1,

$$\phi_{\min}(\tilde{\Omega}_0) \geq \phi_{\min}(\Omega_0) - \|\tilde{\Omega}_0 - \Omega_0\|_2 \geq \phi_{\min}(\Omega_0) - \|\tilde{\Omega}_0 - \Omega_0\|_F \geq \phi_1 - k_1 \phi_1 > 0. \quad (\text{A.7})$$

The first inequality in (A.7) comes from the fact that for any nonzero vector $\boldsymbol{\delta} \in \mathbb{R}^p$, $\boldsymbol{\delta}^T \tilde{\Omega}_0 \boldsymbol{\delta} = \boldsymbol{\delta}^T \Omega_0 \boldsymbol{\delta} + \boldsymbol{\delta}^T (\tilde{\Omega}_0 - \Omega_0) \boldsymbol{\delta} \geq \phi_{\min}(\Omega_0) - \phi_{\max}(\tilde{\Omega}_0 - \Omega_0)$.

Given $\tilde{\Omega}_0 \in \mathcal{S}_+^p \cap \mathcal{S}_{\tilde{E}}^p$, define a new convex set:

$$\mathcal{U}_m(\tilde{\Omega}_0) = \{\mathbf{B} - \tilde{\Omega}_0 \mid \mathbf{B} \in \mathcal{S}_+^p \cap \mathcal{S}_{\tilde{E}}^p\} \subset \mathcal{S}_{\tilde{E}}^p.$$

Let

$$Q(\Omega) = \text{trace}(\Omega \hat{\Sigma}) - \text{trace}(\tilde{\Omega}_0 \hat{\Sigma}) - \log \det \Omega + \log \det \tilde{\Omega}_0.$$

Since the estimate $\hat{\Omega}$ minimizes $Q(\Omega)$, $\hat{\Delta} = \hat{\Omega} - \tilde{\Omega}_0$ minimizes $G(\Delta) = Q(\Delta + \tilde{\Omega}_0)$.

The main idea of this proof is as follows. For a sufficiently large $M > 0$, consider sets

$$\mathcal{T}_1 = \{\Delta \in \mathcal{U}_m(\tilde{\Omega}_0), \|\Delta\|_F = Mr_m\}, \quad \mathcal{T}_2 = \{\Delta \in \mathcal{U}_m(\tilde{\Omega}_0), \|\Delta\|_F \leq Mr_m\},$$

where

$$r_m = \sqrt{\frac{S_0 \log(p-rp)}{m}}.$$

Note that \mathcal{T}_1 is non-empty. Indeed, consider $\mathbf{B}_\epsilon = \epsilon \tilde{\Omega}_0$ for $\epsilon = Mr_m / \|\tilde{\Omega}_0\|_F$. Then $\mathbf{B}_\epsilon = (1 + \epsilon) \tilde{\Omega}_0 - \tilde{\Omega}_0 \in \mathcal{U}_m(\tilde{\Omega}_0)$, hence $\mathbf{B}_\epsilon \in \mathcal{T}_1$. Denote by $\bar{0}$ the matrix of all zero entries. It

is clear that $G(\Delta)$ is convex, and $G(\hat{\Delta}) \leq G(\bar{0}) = Q(\tilde{\Omega}_0) = 0$. Thus if we can show that $G(\Delta) > 0$ for all $\Delta \in \mathcal{T}_1$, the minimizer $\hat{\Delta}$ must be inside \mathcal{T}_2 and hence $\|\hat{\Delta}\|_F \leq Mr_m$. To see this, note that the convexity of $Q(\Omega)$ implies that

$$\inf_{\|\Delta\|_F = Mr_m} Q(\tilde{\Omega}_0 + \Delta) > Q(\tilde{\Omega}_0) = 0.$$

There exists therefore a local minimizer in the ball $\{\tilde{\Omega}_0 + \Delta : \|\Delta\|_F \leq Mr_m\}$, or equivalently, for $\hat{\Delta} \in \mathcal{T}_2$, i.e. $\|\hat{\Delta}\|_F \leq Mr_m$.

In the remainder of the proof, we focus on

$$G(\Delta) = Q(\Delta + \tilde{\Omega}_0) = \text{trace}(\Delta \hat{\Sigma}) - \log \det(\Delta + \tilde{\Omega}_0) + \log \det \tilde{\Omega}_0. \quad (\text{A.8})$$

Applying a Taylor expansion to $\log \det(\tilde{\Omega}_0 + \Delta)$ in (A.8) gives

$$\begin{aligned} & \log \det(\tilde{\Omega}_0 + \Delta) - \log \det \tilde{\Omega}_0 \\ &= \frac{d}{dt} \log \det(\tilde{\Omega}_0 + t\Delta) \Big|_{t=0} \Delta + \int_0^1 (1-t) \frac{d^2}{dt^2} \log \det(\tilde{\Omega}_0 + t\Delta) dt \\ &= \text{trace}(\Delta \tilde{\Sigma}_0) - \text{vec}(\Delta)^T \left\{ \int_0^1 (1-t) (\tilde{\Omega}_0 + t\Delta)^{-1} \otimes (\tilde{\Omega}_0 + t\Delta)^{-1} dt \right\} \text{vec}(\Delta), \quad (\text{A.9}) \end{aligned}$$

where $\text{vec}(\Delta)$ denotes the vectorized Δ , and \otimes is the Kronecker product. For $\Delta \in \mathcal{T}_1$, let K_1 be the integral term in (A.9), and define

$$K_2 = \text{trace} \left\{ \Delta(\hat{\Sigma} - \Sigma_0) \right\}, \quad K_3 = \text{trace} \left\{ \Delta(\tilde{\Sigma}_0 - \Sigma_0) \right\}.$$

We can then write

$$G(\Delta) = K_1 + \text{trace}(\Delta \hat{\Sigma}) - \text{trace}(\Delta \tilde{\Sigma}_0) = K_1 + K_2 - K_3.$$

Next, we bound each of the terms K_1 , K_2 and K_3 to find a lower bound for $G(\Delta)$.

First consider K_2 . Since the diagonal elements of $\hat{\Sigma}$ and Σ_0 are the same after scaling,

$$|K_2| \leq \left| \sum_{i \neq i'} (\hat{\Sigma}_{ii'} - \Sigma_{0,ii'}) \Delta_{ii'} \right|.$$

By Lemma A.3 of Bickel and Levina (2008), there exists a positive constant c_2 depending on $\phi_{\max}(\Sigma_0)$ such that

$$\max_{i \neq i'} |\hat{\Sigma}_{ii'} - \Sigma_{0,ii'}| \leq c_2 \sqrt{\frac{\log(p-rp)}{m}},$$

with probability tending to 1. Let $\Delta^+ = \text{diag}(\Delta)$ be the diagonal matrix with the same diagonal as Δ , and write $\Delta^- = \Delta - \Delta^+$. Then, K_2 is bounded by

$$|K_2| \leq c_2 \sqrt{\frac{\log(p-rp)}{m}} \|\Delta^-\|_1. \quad (\text{A.10})$$

For K_3 , we can use the upper bound for $\|\tilde{\Omega}_0 - \Omega_0\|_F$ in (A.2), and the lower bound for $\phi_{\min}(\tilde{\Omega}_0)$ in (A.7), to write,

$$|K_3| \leq \|\Delta\|_F \|\tilde{\Sigma}_0 - \Sigma_0\|_F \leq \|\Delta\|_F \frac{\|\tilde{\Omega}_0 - \Omega_0\|_F}{\phi_{\min}(\tilde{\Omega}_0) \phi_{\min}(\Omega_0)} \quad (\text{A.11})$$

$$\leq \|\Delta\|_F \frac{c_3 \{S_0 \log(p-rp)/m\}^{1/2}}{(1-k_1)\phi_1^2}. \quad (\text{A.12})$$

The second inequality in (A.11) comes from the rotation invariant property of Frobenius norm, i.e.

$$\|\tilde{\Sigma}_0 - \Sigma_0\|_F = \|\Sigma_0(\Omega_0 - \tilde{\Omega}_0)\tilde{\Sigma}_0\|_F \leq \phi_{\max}(\Sigma_0)\|\Omega_0 - \tilde{\Omega}_0\|_F\phi_{\max}(\tilde{\Sigma}_0).$$

Using (A.2), we can also obtain an upper bound for the maximum eigenvalue of $\tilde{\Omega}_0$:

$$\phi_{\max}(\tilde{\Omega}_0) \leq \phi_{\max}(\Omega_0) + \|\tilde{\Omega}_0 - \Omega_0\|_2 \leq \phi_{\max}(\Omega_0) + \|\tilde{\Omega}_0 - \Omega_0\|_F \leq \frac{1}{\phi_2} + k_1\phi_1.$$

Since $r_m \rightarrow 0$, there exists a sufficiently large $k_2 > 0$ such that for $\Delta \in \mathcal{T}_1$,

$$\|\Delta\|_2 \leq \|\Delta\|_F = Mr_m < \frac{1}{\phi_2}k_2.$$

Following (Rothman *et al.*, 2008, Page 502, proof of Theorem 1), a lower bound for K_1 can be found as

$$\begin{aligned} K_1 &\geq \|\Delta\|_F^2 / \{2(\phi_{\max}(\tilde{\Omega}_0) + \|\Delta\|_2)^2\} \\ &\geq \|\Delta\|_F^2 / \{2(1/\phi_2 + k_1\phi_1 + k_2/\phi_2)^2\} = \frac{\phi_2^2}{2(1 + k_1\phi_1\phi_2 + k_2)^2} \|\Delta\|_F^2. \end{aligned} \quad (\text{A.13})$$

Combining (A.10), (A.12) and (A.13),

$$\begin{aligned} G(\Delta) &\geq \frac{\phi_2^2}{2(1 + k_1\phi_1\phi_2 + k_2)^2} \|\Delta\|_F^2 - c_2 \sqrt{\frac{\log(p - rp)}{m}} \|\Delta^-\|_1 \\ &\quad - \frac{c_3}{(1 - k_1)\phi_1^2} \sqrt{\frac{S_0 \log(p - rp)}{m}} \|\Delta\|_F. \end{aligned}$$

For $\Delta \in \mathcal{T}_1$, applying Cauchy-Schwarz inequality yields

$$\|\Delta^-\|_1 \leq \sqrt{|\hat{E}|} \cdot \|\Delta^-\|_F.$$

We thus have

$$\begin{aligned} G(\Delta) &\geq \frac{\phi_2^2}{2(1 + k_1\phi_1\phi_2 + k_2)^2} \|\Delta\|_F^2 - c_2 \sqrt{\frac{|\hat{E}| \log(p - rp)}{m}} \|\Delta^-\|_F \\ &\quad - \frac{c_3}{(1 - k_1)\phi_1^2} \sqrt{\frac{S_0 \log(p - rp)}{m}} \|\Delta\|_F \\ &\geq \|\Delta\|_F^2 \left\{ \frac{\phi_2^2}{2(1 + k_1\phi_1\phi_2 + k_2)^2} - \frac{c_2}{M} \sqrt{\frac{|\hat{E}|}{S_0}} - \frac{c_3}{M(1 - k_1)\phi_1^2} \right\} > 0, \end{aligned}$$

for M sufficiently large. □

Proof of Corollary 1. Under the assumptions in Theorem 2.2, we have

$$\|\Delta_{\Omega_0}\|_2 = \|\hat{\Omega} - \Omega_0\|_2 = O_{\mathbb{P}} \left(\sqrt{\frac{S_0 \log(p - rp)}{m}} \right) = o_{\mathbb{P}}(1).$$

The partial correlation matrix corresponding to $\hat{\Omega}$ can be written as

$$\hat{\mathbf{A}} = \mathbf{I}_p - \hat{\mathbf{D}}^{-1/2} \hat{\Omega} \hat{\mathbf{D}}^{-1/2} = \mathbf{A}_0 + \mathbf{D}_0^{-1/2} \Omega_0 \mathbf{D}_0^{-1/2} - (\hat{\mathbf{D}})^{-1/2} \hat{\Omega} \hat{\mathbf{D}}^{-1/2} = \mathbf{A}_0 + \Delta_{\mathbf{A}_0},$$

where

$$\begin{aligned} \Delta_{\mathbf{A}_0} &= \mathbf{D}_0^{-1/2} \Omega_0 \mathbf{D}_0^{-1/2} - (\hat{\mathbf{D}})^{-1/2} \hat{\Omega} \hat{\mathbf{D}}^{-1/2} \\ &= \mathbf{D}_0^{-1/2} (\Omega_0 - \hat{\Omega}) \mathbf{D}_0^{-1/2} + \mathbf{D}_0^{-1/2} \hat{\Omega} (\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}) + (\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}) \hat{\Omega} \hat{\mathbf{D}}^{-1/2}. \end{aligned} \quad (\text{A.14})$$

Next we show that each of the summands on the right hand side of (A.14) has ℓ_2 norm $o_{\mathbb{P}}(1)$ and conclude thus $\|\Delta_{\mathbf{A}_0}\|_2 = o_{\mathbb{P}}(1)$.

By Assumption 1, the diagonal entries of Ω_0 satisfy $\omega_{0,ii} \geq \phi_{\min}(\Omega_0) \geq \phi_1$ for all $i = 1, \dots, p$. Thus, $\|\mathbf{D}_0^{-1/2}\|_2 = \max_i \omega_{0,ii}^{-1/2} \leq \phi_1^{-1/2}$. It follows that

$$\|\mathbf{D}_0^{-1/2} (\Omega_0 - \hat{\Omega}) \mathbf{D}_0^{-1/2}\|_2 \leq \|\mathbf{D}_0^{-1/2}\|_2^2 \|\Omega_0 - \hat{\Omega}\|_2 = o_{\mathbb{P}}(1).$$

For the remaining two terms, first notice that $\|\mathbf{D}_0 - \hat{\mathbf{D}}\|_2 \leq \|\mathbf{D}_0 - \hat{\mathbf{D}}\|_F \leq \|\Omega_0 - \hat{\Omega}\|_F = o_{\mathbb{P}}(1)$. Therefore,

$$\begin{aligned} \|\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}\|_2 &= \max_{i=1,\dots,p} |\omega_{0,ii}^{-1/2} - \hat{\omega}_{ii}^{-1/2}| = \max_{i=1,\dots,p} \left| \frac{\omega_{0,ii}^{1/2} - \hat{\omega}_{ii}^{1/2}}{\omega_{0,ii}^{1/2} \hat{\omega}_{ii}^{1/2}} \right| \\ &= \max_{i=1,\dots,p} \left| \frac{\omega_{0,ii} - \hat{\omega}_{ii}}{\omega_{0,ii}^{1/2} \hat{\omega}_{ii}^{1/2} (\omega_{0,ii}^{1/2} + \hat{\omega}_{ii}^{1/2})} \right| \leq \phi_1^{-1} (\phi_1 - o_{\mathbb{P}}(1))^{-1/2} \|\mathbf{D}_0 - \hat{\mathbf{D}}\|_2, \end{aligned}$$

where the last inequality comes from that fact that

$$\min_i |\hat{\omega}_{ii}| = \min_i |\hat{\omega}_{ii} - \omega_{0,ii} + \omega_{0,ii}| \geq \min_i |\omega_{0,ii}| - \max_i |\hat{\omega}_{ii} - \omega_{0,ii}| \geq \phi_1 - o_{\mathbb{P}}(1).$$

Hence, $\|\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}\|_2 = o_{\mathbb{P}}(1)$. Note further,

$$\|\hat{\Omega}\|_2 = \|\hat{\Omega} - \Omega_0 + \Omega_0\|_2 \leq \|\Omega_0\|_2 + \|\hat{\Omega} - \Omega_0\|_2 = \|\Omega_0\|_2 + o_{\mathbb{P}}(1)$$

is bounded above. It follows thus,

$$\begin{aligned} \|\mathbf{D}_0^{-1/2} \hat{\Omega} (\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2})\|_2 &\leq \|\mathbf{D}_0^{-1/2}\|_2 \|\hat{\Omega}\|_2 \|\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}\|_2 = o_{\mathbb{P}}(1), \\ \|(\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}) \hat{\Omega} \hat{\mathbf{D}}^{-1/2}\|_2 &\leq \|\mathbf{D}_0^{-1/2} - \hat{\mathbf{D}}^{-1/2}\|_2 \|\hat{\Omega}\|_2 \|\hat{\mathbf{D}}^{-1/2}\|_2 = o_{\mathbb{P}}(1). \end{aligned}$$

This completes the proof. □

The following proof of Theorem 3.1 adapts from that of Theorem 2.1 in Shojaie and Michailidis (2010).

Proof of Theorem 3.1. Consider the special case where the row vector $\mathbf{b} = \mathbf{1}^T$, i.e. the whole network is tested as one pathway. The general case when $\mathbf{b} \neq \mathbf{1}^T$ follows from a similar argument.

For the partial correlation $\mathbf{A}_0^{(k)}$ ($k = 1, 2$) defined in Section 3.2, it holds that $\Lambda^{(k)}(\Lambda^{(k)})^T = (\mathbf{I}_p - \mathbf{A}_0^{(k)})^{-1} = \sum_{t=0}^{\infty} (\mathbf{A}_0^{(k)})^t$. Hence

$$\begin{aligned}\hat{\Lambda}^{(k)}(\hat{\Lambda}^{(k)})^T &= \sum_{t=0}^{\infty} (\hat{\mathbf{A}}^{(k)})^t = \sum_{t=0}^{\infty} (\mathbf{A}_0^{(k)})^t + \sum_{t=1}^{\infty} \sum_{u=1}^t \binom{t}{u} (\mathbf{A}_0^{(k)})^{t-u} (\Delta_{\mathbf{A}_0^{(k)}})^u \\ &= \Lambda^{(k)}(\Lambda^{(k)})^T + \Delta_{\Lambda^{(k)}}.\end{aligned}$$

For $\hat{\mathbf{A}}^{(k)}$ defined under the assumptions in Theorem 2.2 and 3.1, we have $\|\Delta_{\mathbf{A}_0^{(k)}}\|_2 = o_{\mathbb{P}}(1)$ by Corollary 1. Thus, $\|\Delta_{\Lambda^{(k)}}\|_2 = o_{\mathbb{P}}(1)$.

Using results from Shojaie and Michailidis (2010), the test statistic in (3.11) can be written as

$$TS = \frac{\mathbf{b}(\bar{\mathbf{Y}}^{(2)} - \bar{\mathbf{Y}}^{(1)})}{\sqrt{\hat{\sigma}_{\gamma}^2 \left[\mathbf{b} \left\{ \frac{1}{n_1} \hat{\Lambda}^{(1)}(\hat{\Lambda}^{(1)})^T + \frac{1}{n_2} \hat{\Lambda}^{(2)}(\hat{\Lambda}^{(2)})^T \right\} \mathbf{b}^T \right] + \hat{\sigma}_{\varepsilon}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{b} \mathbf{b}^T}},$$

where $\bar{\mathbf{Y}}^{(k)}$ is the mean expression of genes in the experimental condition k . Shojaie and Michailidis (2010) show that TS is an asymptotically most powerful unbiased test for (3.10) when the correct network information is provided. Therefore, to establish the result in Theorem 3.1, it suffices to show that the denominator of TS is a consistent estimator.

In the following, we first consider the log-likelihood $l_F(\vartheta; \hat{\Lambda})$ based on the estimated networks $\hat{\Lambda} = (\hat{\Lambda}^{(1)}, \hat{\Lambda}^{(2)})$ and correct variance components $\vartheta = (\sigma_{\gamma}^2, \sigma_{\varepsilon}^2)$. We then establish that the maximum likelihood estimator $\hat{\vartheta}_{\hat{\Lambda}} \rightarrow_{\mathbb{P}} \vartheta$ as $\hat{\Lambda}^{(k)}(\hat{\Lambda}^{(k)})^T \rightarrow_{\mathbb{P}} \Lambda^{(k)}(\Lambda^{(k)})^T$ for both k . Hence the denominator of TS is consistent and TS is an asymptotically most powerful unbiased test for (3.10).

Let $\hat{\mathbf{W}}^{(k)} = \sigma_{\gamma}^2 \hat{\Lambda}^{(k)}(\hat{\Lambda}^{(k)})^T + \sigma_{\varepsilon}^2 \mathbf{I}_p$ for $k = 1, 2$. Up to a constant, the negative log-likelihood

$$l_F(\vartheta; \hat{\Lambda}) = \frac{n_1}{2n} l(\vartheta; \hat{\Lambda}^{(1)}) + \frac{n_2}{2n} l(\vartheta; \hat{\Lambda}^{(2)})$$

with

$$\begin{aligned}l(\vartheta; \hat{\Lambda}^{(1)}) &= \log \det(\hat{\mathbf{W}}^{(1)}) + \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbf{R}_j^T (\hat{\mathbf{W}}^{(1)})^{-1} \mathbf{R}_j, \\ l(\vartheta; \hat{\Lambda}^{(2)}) &= \log \det(\hat{\mathbf{W}}^{(2)}) + \frac{1}{n_2} \sum_{j=1+n_1}^n \mathbf{R}_j^T (\hat{\mathbf{W}}^{(2)})^{-1} \mathbf{R}_j,\end{aligned}$$

where $\mathbf{R}_j = \mathbf{Y}_j^{(1)} - \bar{\mathbf{Y}}^{(1)}$ ($j = 1, \dots, n_1$) and $\mathbf{R}_j = \mathbf{Y}_j^{(2)} - \bar{\mathbf{Y}}^{(2)}$ ($j = 1 + n_1, \dots, n$). We treat $l(\vartheta; \hat{\Lambda}^{(1)})$ first. In particular, we can approximate $l(\vartheta; \hat{\Lambda}^{(1)})$ using its one-term Taylor expansion around $\mathbf{W}^{(1)}$

$$l(\vartheta; \hat{\Lambda}^{(1)}) = l(\vartheta; \Lambda^{(1)}) + \text{trace} \{ \nabla_{\mathbf{W}^{(1)}} l(\vartheta; \Lambda^{(1)})^T \Delta_{\mathbf{W}^{(1)}} \} + o(\|\Delta_{\mathbf{W}^{(1)}}\|_2^2),$$

where $\nabla_{\mathbf{W}^{(1)}} l(\vartheta; \Lambda^{(1)})$ is the gradient of $l(\vartheta; \Lambda^{(1)})$ with respect to $\mathbf{W}^{(1)}$ and

$$\nabla_{\mathbf{W}^{(1)}} l(\vartheta; \Lambda^{(1)}) = (\mathbf{W}^{(1)})^{-1} - n_1^{-1} \sum_{j=1}^{n_1} (\mathbf{W}^{(1)})^{-1} \mathbf{R}_j \mathbf{R}_j^T (\mathbf{W}^{(1)})^{-1}.$$

Let $\Gamma = \Delta_{\mathbf{W}^{(1)}} / \|\Delta_{\mathbf{W}^{(1)}}\|_2$ and denote

$$g(\vartheta) = \text{trace} \{ \nabla_{\mathbf{W}^{(1)}} l(\vartheta; \Lambda^{(1)})^T \Gamma \} = \text{trace} \{ (\mathbf{W}^{(1)})^{-1} \Gamma \} - n_1^{-1} \sum_{j=1}^{n_1} \mathbf{R}_j^T (\mathbf{W}^{(1)})^{-1} \Gamma (\mathbf{W}^{(1)})^{-1} \mathbf{R}_j.$$

then

$$l(\vartheta; \hat{\Lambda}^{(1)}) = l(\vartheta; \Lambda^{(1)}) + g(\vartheta) \|\Delta_{\mathbf{W}^{(1)}}\|_2 + o(\|\Delta_{\mathbf{W}^{(1)}}\|_2^2).$$

Using von Neumann's trace inequality (Mirsky, 1975), we can bound the first term in $g(\vartheta)$ by

$$\begin{aligned} |\text{trace} \{ (\mathbf{W}^{(1)})^{-1} \Gamma \}| &\leq \sum_{i=1}^p \varsigma_{[i]}((\mathbf{W}^{(1)})^{-1}) \varsigma_{[i]}(\Gamma) \\ &\leq p \varsigma_{[1]}((\sigma_\gamma^2 \Lambda^{(1)} (\Lambda^{(1)})^T + \sigma_\epsilon^2 \mathbf{I}_p)^{-1}) \varsigma_{[1]}(\Gamma) \\ &= p \frac{1}{\phi_{\min}(\sigma_\gamma^2 \Lambda^{(1)} (\Lambda^{(1)})^T + \sigma_\epsilon^2 \mathbf{I}_p)} \varsigma_{[1]}(\Gamma), \end{aligned}$$

where $\varsigma_{[i]}(\mathbf{A})$ denotes the i th largest singular value of \mathbf{A} . By construction, $\varsigma_{[1]}(\Gamma) = 1$ and $\phi_{\min}(\sigma_\gamma^2 \Lambda^{(1)} (\Lambda^{(1)})^T + \sigma_\epsilon^2 \mathbf{I}_p) \geq \sigma_\epsilon^2$. Hence $|\text{trace} \{ (\mathbf{W}^{(1)})^{-1} \Gamma \}| \leq p/\sigma_\epsilon^2$. On the other hand, with probability tending to 1,

$$\begin{aligned} n_1^{-1} \sum_{j=1}^{n_1} \mathbf{R}_j^T (\mathbf{W}^{(1)})^{-1} \Gamma (\mathbf{W}^{(1)})^{-1} \mathbf{R}_j &\leq \|(\mathbf{W}^{(1)})^{-1} \Gamma (\mathbf{W}^{(1)})^{-1}\|_2 n_1^{-1} \sum_{j=1}^{n_1} \mathbf{R}_j^T \mathbf{R}_j \\ &\leq \|(\mathbf{W}^{(1)})^{-1}\|_2^2 \|\Gamma\|_2 n_1^{-1} \sum_{j=1}^{n_1} \mathbf{R}_j^T \mathbf{R}_j = \sigma_\epsilon^{-4} \mathbb{E}(\|\mathbf{R}_j\|_2^2), \end{aligned}$$

where the last step follows from the strong law of large numbers. This implies that $g(\vartheta)$ is bounded for nontrivial σ_ϵ^2 . Note also $\Delta_{\mathbf{W}^{(1)}} = \hat{\mathbf{W}}^{(1)} - \mathbf{W}^{(1)} = \sigma_\gamma^2 \{ \hat{\Lambda}^{(1)} (\hat{\Lambda}^{(1)})^T - \Lambda^{(1)} (\Lambda^{(1)})^T \} = \sigma_\gamma^2 \Delta_{\Lambda^{(k)}}$. Hence $g(\vartheta) \|\Delta_{\mathbf{W}^{(1)}}\|_2 = g(\vartheta) \sigma_\gamma^2 \|\Delta_{\Lambda^{(k)}}\|_2 = o_{\mathbb{P}}(1)$. Therefore $l(\vartheta; \hat{\Lambda}^{(1)}) = l(\vartheta; \Lambda^{(1)}) + o_{\mathbb{P}}(1)$, and similarly one can show that $l(\vartheta; \hat{\Lambda}^{(2)}) = l(\vartheta; \Lambda^{(2)}) + o_{\mathbb{P}}(1)$. They together imply that

$$l_F(\vartheta; \hat{\Lambda}) = l_F(\vartheta; \Lambda) + o_{\mathbb{P}}(1).$$

Now conditioning on the event $\{l_F(\vartheta; \hat{\Lambda}) = l_F(\vartheta; \Lambda)\}$, the estimate of the variance components is $\hat{\vartheta} = \arg\min_{\vartheta} l_F(\vartheta; \Lambda)$. Since $l_F(\vartheta; \Lambda)$ is convex with respect to ϑ , M-estimation results in Haberman (1989) imply that $\mathbb{P}(\hat{\vartheta} = \vartheta) = 1$ and hence $\hat{\vartheta} \rightarrow_{\mathbb{P}} \vartheta$ as $\hat{\Lambda}^{(k)} (\hat{\Lambda}^{(k)})^T \rightarrow_{\mathbb{P}} \Lambda^{(k)} (\Lambda^{(k)})^T$ for both k . It follows immediately that the denominator of the test statistic TS is a consistent estimator as $\hat{\Lambda}^{(k)} (\hat{\Lambda}^{(k)})^T \rightarrow_{\mathbb{P}} \Lambda^{(k)} (\Lambda^{(k)})^T$ for both k . This concludes the proof. \square

B Efficient Estimation of Model Parameters

In this section, we present in details the strategy used to scale up the NetGSA algorithm for large scale networks as well as necessary derivations.

As pointed out in Section 2.2.1 of the main text, inference in NetGSA requires estimation of the mean parameters $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ and variance components σ_γ^2 and σ_ϵ^2 . After rearranging the data \mathcal{D} to be a $N \times 1$ vector \mathbf{Y} , we can write the model using the matrix notation as

$$\mathbf{Y} = \Psi \boldsymbol{\beta} + \Pi \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{B.1}$$

where the design matrix

$$\Pi = \text{bdiag}(\Lambda^{(1)}, \dots, \Lambda^{(1)}, \Lambda^{(2)}, \dots, \Lambda^{(2)}) \in \mathbf{R}^{N \times N},$$

and

$$\Psi = \begin{pmatrix} \Lambda^{(1)} & & \\ \vdots & & \\ \Lambda^{(1)} & & \\ & \Lambda^{(2)} & \\ & \vdots & \\ & \Lambda^{(2)} & \end{pmatrix} \in \mathbf{R}^{N \times 2p}.$$

The variance of \mathbf{Y} , i.e. $\mathbf{W} = \sigma_\epsilon^2 \mathbf{I}_N + \sigma_\gamma^2 \Pi \Pi'$. The mean β can be estimated via the maximum likelihood as

$$\hat{\beta} = (\Psi' \hat{\mathbf{W}}^{-1} \Psi)^{-1} \Psi' \hat{\mathbf{W}}^{-1} \mathbf{Y},$$

where $\hat{\mathbf{W}}$ is defined using the estimated variances. The variances are often estimated via the maximum likelihood or restricted maximum likelihood using the profile likelihood. Thus, one can use an iterative algorithm to jointly estimate β and the variance components.

However, estimation of the variance components is computationally demanding for large networks. To ensure stability, the earlier version of the NetGSA considered profiling out one of the variance components and implemented an algorithm from Byrd *et al.* (1995), which uses a limited-memory modification of the Broyden–Fletcher–Goldfarb–Shanno quasi-Newton method to optimize the profile log-likelihood. However, the above implementation has a few issues. The first issue is its high computational cost due to the inefficient evaluation of matrix inverses and determinants. Moreover, the algorithm from Byrd *et al.* (1995) requires finite values of the objective function within the supplied box constraints, which is often not satisfied, even after the constraints are adjusted to be within a small range of the optimal estimate. This is particularly the case when the underlying networks are large. To extend the applicability of the NetGSA, we consider using Newton’s method for estimating the variance parameters based on the profile log-likelihood to improve the computational stability. In particular, we make the following two key improvements for implementation of Newton’s method.

First, it is clear that $\text{Var}(\mathbf{Y}_j^{(k)}) = \sigma_\epsilon^2 \{ \mathbf{I}_p + \tau \Lambda^{(k)} (\Lambda^{(k)})^T \} = \sigma_\epsilon^2 \Sigma^{(k)}$, where $\tau = \sigma_\gamma^2 / \sigma_\epsilon^2$. Since the profile log-likelihood as well as its gradient and Hessian matrix with respect to τ all depend on $\Sigma^{(k)}$ ($k = 1, 2$) and their inverses, we choose to invert from their Cholesky decompositions $\Sigma^{(k)} = \mathbf{U}^T \mathbf{U}$, where \mathbf{U} is an upper triangular matrix. The inversion of the triangular matrices results in significant speedup and the inverses of the original matrices can then be computed as $(\Sigma^{(k)})^{-1} = (\mathbf{U}^{-1})(\mathbf{U}^{-1})^T$. In the meantime, we also simplify the calculation of the determinant of $\Sigma^{(k)}$ since $\det(\Sigma^{(k)}) = \det(\mathbf{U})^2$, which is necessary for evaluating the profile log-likelihood.

Second, the quality of the starting point as well as step sizes will both affect convergence of Newton’s method. To select a good starting point, we use a method-of-moment-type estimate of the variance components. Specifically, denote the residuals $\mathbf{R}_j = \mathbf{Y}_j^{(k)} - \Lambda^{(k)} \hat{\boldsymbol{\mu}}^{(k)}$ for $j = 1, \dots, n$, where $\hat{\boldsymbol{\mu}}^{(k)}$ is the estimate of $\boldsymbol{\mu}^{(k)}$. Assume that there is a single variance σ_ϵ^2 that applies to all ϵ_j ($j = 1, \dots, n$) and variances of γ_j are different. The variance of \mathbf{R}_j can be decomposed as $(\sigma_\gamma^2)_j + \sigma_\epsilon^2$. We then take the minimum of $\text{Var}(\mathbf{R}_j)$ as the estimate of σ_ϵ^2 and average of the remaining variances as the estimate of σ_γ^2 . Their ratio is used as the initial value for τ . The approximation runs very fast and does not add much computational cost to

the method. To find the appropriate step sizes, we use backtracking line search as described in (Boyd and Vandenberghe, 2004, page 464).

With the above two modifications, Newton's method can then be implemented to optimize the profile log-likelihood and returns an estimate of τ . Estimates of $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\epsilon^2$ follow immediately. The implementation of Newton's method requires the gradient and the Hessian of the objective function, i.e., the profile log-likelihood. Next we provide details on how to calculate these quantities from the profile log-likelihood when profiling out σ_ϵ , based on the general framework introduced in Lindstrom and Bates (1988). The derivation follows similarly when profiling out σ_γ .

Let $N = np$ be the total number of observations for all genes. Recall that for $k = 1, 2$, $\Sigma^{(k)} = \mathbf{I}_p + \tau \Lambda^{(k)} (\Lambda^{(k)})^T$ with $\tau = \sigma_\gamma^2 / \sigma_\epsilon^2$. The residuals $\mathbf{R}_j = \mathbf{Y}_j^{(k)} - \Lambda^{(k)} \hat{\boldsymbol{\mu}}^{(k)}$ for $j = 1, \dots, n$, where $\hat{\boldsymbol{\mu}}^{(k)}$ is the estimate of $\boldsymbol{\mu}^{(k)}$. Given the observations $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ (with the first n_1 samples from condition 1 and the remaining $n_2 = n - n_1$ samples from condition 2), the nonconstant part of the "full" log-likelihood l_F is

$$l_F(\sigma_\epsilon, \tau \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n) = -\frac{1}{2} \left\{ n_1 \log \det(\sigma_\epsilon^2 \Sigma^{(1)}) + n_2 \log \det(\sigma_\epsilon^2 \Sigma^{(2)}) \right\} \\ - \frac{1}{2} \sigma_\epsilon^{-2} \left\{ \sum_{j=1}^{n_1} \mathbf{R}_j^T (\Sigma^{(1)})^{-1} \mathbf{R}_j + \sum_{j=n_1+1}^n \mathbf{R}_j^T (\Sigma^{(2)})^{-1} \mathbf{R}_j \right\}.$$

Similarly, the nonconstant part of the log-likelihood using the restricted maximum likelihood is

$$l_R(\sigma_\epsilon, \tau \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n) = l_F(\sigma_\epsilon, \tau \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n) \\ - \frac{1}{2} \log \det \left\{ n_1 \sigma_\epsilon^{-2} (\Lambda^{(1)})^T (\Sigma^{(1)})^{-1} \Lambda^{(1)} + n_2 \sigma_\epsilon^{-2} (\Lambda^{(2)})^T (\Sigma^{(2)})^{-1} \Lambda^{(2)} \right\}.$$

We first solve for σ_ϵ^2 as a function of τ . The maximum likelihood estimate of σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N} \left\{ \sum_{j=1}^{n_1} \mathbf{R}_j^T (\Sigma^{(1)})^{-1} \mathbf{R}_j + \sum_{j=n_1+1}^n \mathbf{R}_j^T (\Sigma^{(2)})^{-1} \mathbf{R}_j \right\}, \quad (\text{B.2})$$

whereas its restricted maximum likelihood estimate is

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N - 2p} \left\{ \sum_{j=1}^{n_1} \mathbf{R}_j^T (\Sigma^{(1)})^{-1} \mathbf{R}_j + \sum_{j=n_1+1}^n \mathbf{R}_j^T (\Sigma^{(2)})^{-1} \mathbf{R}_j \right\}. \quad (\text{B.3})$$

Substituting σ_ϵ^2 with its corresponding estimate, we obtain the profile log-likelihood

$$p_F(\tau \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n) = -\frac{1}{2} (n_1 \log \det \Sigma^{(1)} + n_2 \log \det \Sigma^{(2)}) \\ - \frac{1}{2} N \log \left\{ \sum_{j=1}^{n_1} \mathbf{R}_j^T (\Sigma^{(1)})^{-1} \mathbf{R}_j + \sum_{j=n_1+1}^n \mathbf{R}_j^T (\Sigma^{(2)})^{-1} \mathbf{R}_j \right\}, \quad (\text{B.4})$$

for maximum likelihood and

$$\begin{aligned}
p_R(\tau \mid \mathbf{Y}_1, \dots, \mathbf{Y}_n) = & -\frac{1}{2}(n_1 \log \det \Sigma^{(1)} + n_2 \log \det \Sigma^{(2)}) \\
& -\frac{1}{2}(N-2p) \log \left\{ \sum_{j=1}^{n_1} \mathbf{R}_j^T (\Sigma^{(1)})^{-1} \mathbf{R}_j + \sum_{j=n_1+1}^n \mathbf{R}_j^T (\Sigma^{(2)})^{-1} \mathbf{R}_j \right\} \\
& -\frac{1}{2} \log \det \left\{ n_1 (\Lambda^{(1)})^T (\Sigma^{(1)})^{-1} \Lambda^{(1)} + n_2 (\Lambda^{(2)})^T (\Sigma^{(2)})^{-1} \Lambda^{(2)} \right\},
\end{aligned} \tag{B.5}$$

for restricted maximum likelihood.

As $\Sigma^{(k)}$ ($k = 1, 2$) are the only terms that depend on τ , we first look at the derivatives of $\log \det \Sigma^{(k)}$, $\mathbf{R}_j^T (\Sigma^{(k)})^{-1} \mathbf{R}_j$, and $\log \det(\mathbf{H})$ with respect to τ , where $\mathbf{H} = n_1 \mathbf{H}^{(1)} + n_2 \mathbf{H}^{(2)}$ and $\mathbf{H}^{(k)} = (\Lambda^{(k)})^T (\Sigma^{(k)})^{-1} \Lambda^{(k)}$ for $k = 1, 2$. Let

$$\mathbf{B}^{(k)} = (\Sigma^{(k)})^{-1} \frac{d\Sigma^{(k)}}{d\tau} (\Sigma^{(k)})^{-1}.$$

Then

$$\begin{aligned}
\frac{d \log \det(\Sigma^{(k)})}{d\tau} &= \text{trace} \left\{ (\Sigma^{(k)})^{-1} \frac{d\Sigma^{(k)}}{d\tau} \right\}, \\
\frac{d^2 \log \det(\Sigma^{(k)})}{d\tau^2} &= \text{trace} \left\{ -(\mathbf{B}^{(k)})^T \frac{d\Sigma^{(k)}}{d\tau} + (\Sigma^{(k)})^{-1} \frac{d^2 \Sigma^{(k)}}{d\tau^2} \right\}, \\
\frac{d \mathbf{R}_j^T (\Sigma^{(k)})^{-1} \mathbf{R}_j}{d\tau} &= -\mathbf{R}_j^T \mathbf{B}^{(k)} \mathbf{R}_j, \quad \frac{d^2 \mathbf{R}_j^T (\Sigma^{(k)})^{-1} \mathbf{R}_j}{d\tau^2} = -\mathbf{R}_j^T \frac{d\mathbf{B}^{(k)}}{d\tau} \mathbf{R}_j, \\
\frac{d \log \det(\mathbf{H})}{d\tau} &= -\text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k (\Lambda^{(k)})^T \mathbf{B}^{(k)} \Lambda^{(k)} \right\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{d^2 \log \det(\mathbf{H})}{d\tau^2} &= -\text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k (\Lambda^{(k)})^T \mathbf{B}^{(k)} \Lambda^{(k)} \times \mathbf{H}^{-1} \sum_{k=1,2} n_k (\Lambda^{(k)})^T \frac{d\mathbf{B}^{(k)}}{d\tau} \Lambda^{(k)} \right\} \\
&\quad - \text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k (\Lambda^{(k)})^T \frac{d^2 \mathbf{B}^{(k)}}{d\tau^2} \Lambda^{(k)} \right\},
\end{aligned}$$

where

$$\frac{d\mathbf{B}^{(k)}}{d\tau} = -(\Sigma^{(k)})^{-1} \left\{ 2 \frac{d\Sigma^{(k)}}{d\tau} (\Sigma^{(k)})^{-1} \frac{d\Sigma^{(k)}}{d\tau} - \frac{d^2 \Sigma^{(k)}}{d\tau^2} \right\} (\Sigma^{(k)})^{-1}.$$

Given the covariance $\Sigma^{(k)}$ ($k = 1, 2$) defined in Section 3.1, we can further simplify the above derivatives and obtain

$$\frac{d \log \det \Sigma^{(k)}}{d\tau} = \text{trace} \left\{ \mathbf{H}^{(k)} \right\}, \quad \frac{d^2 \log \det \Sigma^{(k)}}{d\tau^2} = -\text{trace} \left\{ \mathbf{H}^{(k)} \mathbf{H}^{(k)} \right\},$$

$$\begin{aligned}
\frac{d \mathbf{R}_j^T (\Sigma^{(k)})^{-1} \mathbf{R}_j}{d\tau} &= -\mathbf{R}_j^T (\Sigma^{(k)})^{-1} \Lambda^{(k)} (\Lambda^{(k)})^T (\Sigma^{(k)})^{-1} \mathbf{R}_j, \\
\frac{d^2 \mathbf{R}_j^T (\Sigma^{(k)})^{-1} \mathbf{R}_j}{d\tau^2} &= 2\mathbf{R}_j^T (\Sigma^{(k)})^{-1} \Lambda^{(k)} \mathbf{H}^{(k)} (\Lambda^{(k)})^T (\Sigma^{(k)})^{-1} \mathbf{R}_j, \\
\frac{d \log \det(\mathbf{H})}{d\tau} &= -\text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k \mathbf{H}^{(k)} \mathbf{H}^{(k)} \right\}, \\
\frac{d^2 \log \det(\mathbf{H})}{d\tau^2} &= -\text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k \mathbf{H}^{(k)} \mathbf{H}^{(k)} \right\} + 2 \text{trace} \left\{ \mathbf{H}^{-1} \sum_{k=1,2} n_k \mathbf{H}^{(k)} \mathbf{H}^{(k)} \mathbf{H}^{(k)} \right\}.
\end{aligned}$$

With the above quantities, one can then calculate the gradient and Hessian of the profile log-likelihood p_R for restricted maximum likelihood and use Newton's method to obtain an estimate of τ . Estimate of $\hat{\sigma}_\epsilon^2$ is calculated from (B.3), and $\hat{\sigma}_\gamma^2 = \hat{\tau} \hat{\sigma}_\epsilon^2$. Estimation with maximum likelihood follows similarly by applying Newton's method to p_F and utilizing (B.2).

C Additional Simulation Results

To benchmark the performance of the proposed network estimation procedure as well as NetGSA, we first revisit the two simulation experiments presented in Section 3 of the main paper and report the Type I error (or the observed false discovery proportion) when the null hypothesis is true. In addition, we consider two other simulation experiments and refer to them as the third and fourth settings, following the earlier two settings in the main paper. The simulations in this section are also discussed when comparing the run time of NetGSA with different variance estimation algorithms in Section 3 of the main paper.

C.1 Simulation Studies 1 and 2

C.1.1 Powers

We have shown the estimated powers in Tables 2 and 3 in the main paper for the two experiments based on adjusted false discovery rate (FDR) cutoffs. For completeness, we present here the estimated powers in Tables A1 and A2 when the FDR cutoff is $q^* = 0.05$. Due to the use of different FDR cutoffs, one expects to see higher powers for the columns corresponding to 0.2, 0.8, 0.2(m) and 0.8(m), and slightly lower powers for E and GSA-c in Tables A1 and A2 compared to, respectively, Tables 2 and 3 in the main paper. In both Table A1 and Table A2, we still observe the following: NetGSA with the exact networks does a very good job in recovering the true powers for each pathway; NetGSA with more external structural information generally reports powers that are closer to the true power; further NetGSA is robust to misspecification in external structural information. Further, for pathway 1 that has neither mean nor structural changes, we note that the powers are sometimes greater than 0.05 when NetGSA with estimated network information is applied. This is partly due to the network estimation error.

Table A1: Powers with false discovery rate cutoff $q^* = 0.05$ in experiment 1. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information.

Pathway	0.2	0.8	E	T	$p = 100$			
					GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.05	0.10	0.03	0.06	0.17	0.01	0.05	0.08
2	0.18	0.13	0.03	0.06	0.09	0.00	0.16	0.16
3	0.50	0.48	0.30	0.46	0.36	0.00	0.50	0.63
4	0.49	0.29	0.02	0.07	0.25	0.04	0.44	0.29
5	0.94	0.98	0.89	0.97	0.97	0.00	0.95	0.95
6	0.46	0.49	0.20	0.26	0.36	0.00	0.49	0.41
7	0.84	0.90	0.94	0.99	0.98	0.04	0.82	0.92
8	0.54	0.68	0.42	0.57	0.87	0.00	0.56	0.61

C.1.2 Type I errors

To validate the type I error when the null hypothesis is true, we use the same null setup as presented in Section 3 of the main paper for both experiment 1 and 2. The network structure and node mean expressions under the alternative are set to be the same as in the null case. We use $n_1 = n_2 = 25$ samples for each condition in experiment 1 and $n_1 = n_2 = 40$ in experiment 2 for pathway enrichment analysis. When the underlying networks are not available, we estimate the networks based on external information ranging from 0%, 20%, 80% to 100% and 100 observations generated from the true network. Scenarios with misspecified structural information are also considered. In the following, we present type I errors based on both the adjusted FDR cutoffs and uniform FDR cutoff at $q^* = 0.05$, where the former corresponds to using $q^* = 0.01$ for cases 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Table A3 and A4 present the type I errors based on adjusted FDR cutoffs evaluated over 100 replications for experiment 1 and 2, respectively. As expected, the type I errors when all true parameters are plugged in the NetGSA model are 0.05 for all subnetworks in both experiments. When the exact networks are known, one only estimates the variance components in the NetGSA model and observes small false discovery proportions. When the exact networks are not available such that one estimates the partial correlations as well as the variance components, the type I errors are generally greater than q^* ; in particular, the type I errors get worse as the amount of external information decreases. This is likely due to the small sample sizes for estimating the networks. In general, one benefits from having more external structural information and/or more observations for recovering the underlying networks when using NetGSA. In comparison, both GSA-c and GSA-s have type I errors smaller than 0.05.

As a comparison, Table A5 and A6 present the type I errors based on the uniform FDR cutoffs 0.05 evaluated over 100 replications for experiment 1 and 2, respectively. The reported false discovery proportions for NetGSA with estimated networks including columns 0.2, 0.8, 0.2(m) and 0.8(m) are generally higher than the corresponding columns in Table A3 and A4, especially pathways 2-4 and 6-8.

It is important to make a distinction between the samples used for enrichment analysis and

Table A2: Powers with false discovery rate cutoff $q^* = 0.05$ in experiment 2. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information.

Pathway	0.2	0.8	E	T	$p = 160$			
					GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.10	0.11	0.02	0.05	0.07	0.01	0.10	0.11
2	0.52	0.60	0.15	0.36	0.51	0.00	0.53	0.60
3	0.96	1.00	0.95	0.99	1.00	0.00	0.97	1.00
4	0.98	1.00	1.00	1.00	1.00	0.09	0.98	1.00
5	0.38	0.34	0.02	0.11	0.10	0.03	0.41	0.36
6	0.46	0.35	0.01	0.07	0.24	0.01	0.46	0.34
7	0.78	0.83	0.89	0.92	0.99	0.00	0.78	0.82
8	0.92	0.99	1.00	1.00	1.00	0.02	0.91	0.98

Table A3: Type I error when the null hypothesis is true in experiment 1. False discovery rate cutoffs are $q^* = 0.01$ for cases 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.00	0.00	0.00	0.05	0.02	0.03	0.00	0.01
2	0.06	0.06	0.01	0.05	0.03	0.05	0.07	0.02
3	0.29	0.16	0.02	0.05	0.01	0.02	0.28	0.14
4	0.16	0.12	0.00	0.05	0.03	0.05	0.17	0.11
5	0.02	0.01	0.01	0.05	0.00	0.03	0.03	0.01
6	0.24	0.12	0.00	0.05	0.01	0.02	0.23	0.12
7	0.18	0.13	0.03	0.05	0.03	0.05	0.19	0.14
8	0.11	0.10	0.00	0.05	0.02	0.02	0.09	0.09

those for network estimation. If one has access to a large number of observations that can only be used for network estimation as well as some external structural information, then NetGSA can leverage both resources to achieve more reliable enrichment testing. However, methods like GSA are unable to take advantage of such rich external information.

C.2 Simulation Studies 3 and 4

C.2.1 The setup

Our third simulation experiment considers an undirected network with $p = 160$ and a design similar to the second experiment. However, in this case, each of the 8 subnetworks has a denser structure and there are more interactions between subnetworks. Specifically, there are 70 edges connecting the 20 nodes in each subnetwork under the null. There is 30% chance of an interaction between four randomly selected nodes from each subnetwork to four randomly selected nodes from every other subnetwork. Under the alternative, there is an increase of

Table A4: Type I error when the null hypothesis is true in experiment 2. False discovery rate cutoffs are $q^* = 0.01$ for cases 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.03	0.03	0.00	0.05	0.01	0.03	0.03	0.03
2	0.08	0.05	0.00	0.05	0.01	0.05	0.08	0.05
3	0.17	0.09	0.00	0.05	0.01	0.01	0.17	0.08
4	0.34	0.25	0.01	0.05	0.03	0.03	0.35	0.25
5	0.25	0.09	0.00	0.05	0.01	0.02	0.27	0.11
6	0.32	0.18	0.00	0.05	0.02	0.04	0.33	0.18
7	0.23	0.13	0.00	0.05	0.00	0.04	0.28	0.13
8	0.26	0.12	0.00	0.05	0.02	0.01	0.29	0.11

Table A5: Type I error when the null hypothesis is true in experiment 1. False discovery rate cutoff is $q^* = 0.05$.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.03	0.04	0.00	0.05	0.01	0.01	0.04	0.03
2	0.18	0.18	0.01	0.05	0.04	0.02	0.18	0.17
3	0.32	0.25	0.00	0.05	0.02	0.00	0.36	0.25
4	0.39	0.29	0.01	0.05	0.03	0.00	0.39	0.32
5	0.09	0.04	0.00	0.05	0.02	0.05	0.11	0.07
6	0.30	0.19	0.00	0.05	0.04	0.03	0.32	0.21
7	0.39	0.29	0.01	0.05	0.04	0.02	0.39	0.32
8	0.29	0.19	0.02	0.05	0.04	0.03	0.27	0.21

0.5 in mean values for varying proportions of nodes (0%, 30% and 50%) for subnetworks 1-3 and 5-7. For subnetworks 4 and 8, 70% of the nodes have mean values decreased by 0.5. Moreover, 13% of the edges in subnetworks 5-8 under the alternative are different from their null counterpart.

The fourth experiment uses an undirected network of size $p = 400$ and illustrates the scalability of the proposed method using the new optimization algorithm. The network consists of 20 subnetworks, each corresponding to a pathway with 20 genes. The probability of an interaction between the hub node in one subnetwork and two randomly selected nodes from another subnetwork is 0.4. Under the null, all subnetworks have the same topology generated from a scale-free random graph such that there are 37 edges linking the 20 nodes; all the nodes have mean expression values 1. Under the alternative, subnetworks 1-6 and 11-16 keep the same mean expression values, but 20%, 40%, 60% and 80% of nodes in subnetworks 7-10 and 17-20 have 0.5 unit increase in their mean values, respectively. In addition, subnetworks 11-20 under the alternative all have 39 edges and their structure differs from their null equivalent by 30%. This experiment is also of interest because we created a setting where there are enough subnetworks in order for the permutation based Gene Set Analysis (Efron and Tibshirani, 2007) to calibrate the number of permutations required.

In both experiments, we also included scenarios where a proportion of the supplied structural information is incorrectly specified. This is to check whether NetGSA is robust to model

Table A6: Type I error when the null hypothesis is true in experiment 2. False discovery rate cutoff is $q^* = 0.05$.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2m	0.8m
1	0.15	0.14	0.00	0.05	0.01	0.03	0.13	0.15
2	0.14	0.12	0.00	0.05	0.00	0.01	0.15	0.12
3	0.25	0.17	0.00	0.05	0.01	0.01	0.26	0.17
4	0.33	0.26	0.00	0.05	0.04	0.01	0.32	0.28
5	0.46	0.32	0.00	0.05	0.01	0.02	0.42	0.32
6	0.43	0.26	0.00	0.05	0.02	0.03	0.44	0.26
7	0.40	0.30	0.00	0.05	0.02	0.02	0.40	0.32
8	0.47	0.29	0.00	0.05	0.01	0.01	0.46	0.29

misspecification. In particular, about 50% (20%) of the supplied edges are actually not present in the true model for the case $r = 0.2$ ($r = 0.8$).

C.2.2 Network estimation

Table A7 presents the deviance measures for estimating the networks with 100 replicates and sample sizes of $m = 500$ for $p = 160$ and $m = 400$ for $p = 400$, when varying levels of external information are available. In both experiments, we see performance improvement in Matthews correlation coefficient and Frobenius norm loss as the correctly specified structural information of the networks r increases ($r = 0.2, 0.8$ corresponding to 20% and 80% total information, respectively). When there exists misspecified edges in the external information (denoted by 0.2(m) and 0.8(m)), we used two tuning parameters for network estimation, one for controlling the overall sparsity of the network and the other for correcting the misspecified edges. The optimal tuning parameters were selected over a grid of values using BIC. It can be seen that the performance of network estimation is not compromised by much after properly selected tuning parameters.

Table A7: Deviance measures for network estimation in experiment 3 and 4. FPR(%), false positive rate in percentage; FNR(%), false negative rate in percentage; MCC, Matthews correlation coefficient; Fnorm, Frobenius norm loss.

		$p = 160$				$p = 400$			
	r	FPR(%)	FNR(%)	MCC	Fnorm	FPR(%)	FNR(%)	MCC	Fnorm
Null	0.0	8.20	12.38	0.59	0.58	4.19	14.58	0.44	0.50
	0.2	7.21	14.88	0.60	0.58	3.89	12.29	0.46	0.48
	0.8	3.04	5.47	0.80	0.47	1.88	3.87	0.65	0.40
	0.2(m)	7.48	12.83	0.60	0.57	3.89	12.44	0.46	0.49
	0.8(m)	3.07	8.39	0.78	0.49	1.88	4.23	0.64	0.40
Alternative	0.0	8.16	11.62	0.59	0.57	4.25	14.70	0.44	0.50
	0.2	7.15	14.41	0.60	0.57	3.96	12.38	0.46	0.48
	0.8	3.02	5.24	0.80	0.46	1.95	3.75	0.64	0.40
	0.2(m)	7.42	12.60	0.60	0.56	3.97	12.57	0.46	0.48
	0.8(m)	3.03	8.17	0.78	0.48	1.95	4.13	0.64	0.40

C.2.3 Powers

Table A8 shows the estimated powers after correcting for false discovery rate in the third experiment with $p = 160$. When the exact networks are known, NetGSA estimated powers match very well with the true powers. In the case of unknown networks, we see consistent recovery of high powers for subnetworks 3, 4, 6, 7 and 8 using NetGSA even with only 20% external information. This suggests that, with large enough samples for network estimation, a small amount of external knowledge is sufficient for making reliable inference using the network-based method. Interestingly, GSA-c identifies only subnetworks 3 and 4 as significantly differential with high power, whereas GSA-s returns relatively high power for subnetworks 3 and 6 but surprisingly low power for subnetwork 8. One possible reason for this pattern is that the busy interactions between subnetworks and the negative mean changes in subnetworks 4 and 8 affected the ability of GSA to properly recognize the correct differential behavior. The last two columns in Table A8 show the estimated powers from NetGSA when the external information is misspecified. For both cases ($r = 20\%$ and $r = 80\%$), the results bear high similarity to those in the first two columns, which suggests that NetGSA is robust to model misspecification.

Table A8: Powers in experiment 3. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m), 0.8(m) and GSA-s, 0.10 for E and GSA-c.

Pathway	$p = 160$							
	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.00	0.00	0.05	0.05	0.06	0.04	0.00	0.00
2	0.49	0.50	0.77	0.71	0.59	0.04	0.49	0.52
3	0.75	0.71	0.91	0.90	0.97	0.76	0.78	0.72
4	0.90	0.88	1.00	0.99	0.59	0.97	0.89	0.87
5	0.18	0.12	0.07	0.05	0.48	0.03	0.17	0.13
6	0.47	0.44	0.66	0.68	0.78	0.27	0.48	0.44
7	0.68	0.62	0.96	0.95	0.64	0.04	0.69	0.60
8	0.76	0.77	0.99	1.00	0.14	0.50	0.77	0.75

The estimated powers after correcting for false discovery rate in the fourth experiment are shown separately in Table A10. When the exact networks with the correct edge weights are known, we again see that NetGSA estimated powers match the true powers closely, with very low powers for subnetworks 1-6 which have no changes in neither mean expressions nor structures, high powers for subnetworks 8-10 which have significant changes in mean expression values, low powers for subnetworks 11-16 that have changes in structures and very high powers for pathways 17-20 with changes in both. When there is 20% external information on the underlying network topology, NetGSA's powers for subnetworks 8-10 and 18-20 are close to true powers. However, NetGSA overestimates the powers for subnetworks 11-16. This is due to the small sample size ($m = 400$) for estimating the underlying networks. When the external information is slightly misspecified, the last two columns indicate that NetGSA still returns valid powers that are comparable to those obtained with correctly specified structural

Table A9: Powers in experiment 3. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information. False discovery rate cutoffs are $q^* = 0.05$.

Pathway	0.2	0.8	E	T	$p = 160$			
					GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.02	0.02	0.01	0.05	0.02	0.00	0.02	0.02
2	0.70	0.72	0.67	0.71	0.65	0.02	0.72	0.75
3	0.81	0.78	0.87	0.90	0.97	0.39	0.80	0.80
4	0.95	0.92	0.99	0.99	0.53	0.79	0.94	0.92
5	0.30	0.32	0.03	0.05	0.44	0.00	0.27	0.31
6	0.56	0.65	0.63	0.68	0.88	0.11	0.57	0.66
7	0.78	0.74	0.93	0.95	0.72	0.06	0.78	0.76
8	0.90	0.87	1.00	1.00	0.12	0.27	0.88	0.88

information. In comparison, GSA-s believes almost all subnetworks except 1-6 are significantly differential. On the other hand, when testing against the competitive null, the results from GSA-c suggest that only subnetworks 10, 19 and 20 are significantly differential. The conflicting results from GSA with or without randomization of the genes also raise concerns as to which version to choose in practice.

C.2.4 Type I errors

Finally, we also look at the scenarios where the null hypothesis is true for experiment 3 and 4. Again we present type I errors obtained based on both the adjusted FDR cutoffs and the uniform FDR cutoffs at 0.05, where the adjusted FDR cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c. The results can be found in Tables A12, A13, A14 and A15.

Since the sample size used for network estimation in experiment 3 is sufficiently large, we observe very good control of type I errors in Table A12 for NetGSA, even with estimated networks. In Table A14, type I errors are high for some pathways, which is again due to the small sample size for estimating 400×400 networks.

D Additional Results on Metabolomics and Genomics

Table A16, A17 and A18 present the full list of pathways used in each of the studies and their corresponding false discovery rate corrected p -values, respectively.

Table A10: Powers in experiment 4. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Pathway	$p = 400$							
	0.2	0.8	E	T	GSA-s	GSA-c	0.2(m)	0.8(m)
1	0.02	0.02	0.05	0.05	0.09	0.03	0.02	0.02
2	0.14	0.12	0.06	0.05	0.13	0.03	0.15	0.12
3	0.19	0.18	0.05	0.05	0.11	0.01	0.18	0.18
4	0.29	0.24	0.03	0.05	0.15	0.04	0.30	0.25
5	0.32	0.29	0.05	0.05	0.15	0.02	0.32	0.29
6	0.46	0.44	0.05	0.05	0.18	0.03	0.44	0.43
7	0.56	0.50	0.85	0.83	0.93	0.00	0.58	0.52
8	0.75	0.73	1.00	1.00	1.00	0.00	0.75	0.73
9	0.92	0.89	1.00	1.00	1.00	0.19	0.93	0.89
10	0.98	0.99	1.00	1.00	1.00	1.00	0.98	0.99
11	0.49	0.46	0.09	0.06	1.00	0.00	0.47	0.48
12	0.57	0.60	0.07	0.07	1.00	0.00	0.54	0.59
13	0.59	0.57	0.07	0.05	1.00	0.05	0.59	0.57
14	0.58	0.63	0.12	0.07	0.99	0.03	0.59	0.63
15	0.58	0.66	0.07	0.07	0.99	0.02	0.57	0.66
16	0.60	0.50	0.08	0.07	1.00	0.03	0.57	0.51
17	0.65	0.68	0.90	0.89	1.00	0.02	0.63	0.68
18	0.73	0.81	1.00	1.00	1.00	0.34	0.74	0.80
19	0.85	0.85	1.00	1.00	1.00	1.00	0.85	0.86
20	0.86	0.88	1.00	1.00	1.00	1.00	0.86	0.87

Table A11: Powers in experiment 4. 0.2/0.8 refer to NetGSA with 20%/80% external information; E refers to NetGSA with the exact networks; T refers to the true power; GSA-c/GSA-s refer to Gene Set Analysis with/without randomization of the genes in 1000 permutations, respectively; 0.2(m)/0.8(m) refer to NetGSA with 20%/80% misspecified external information. False discovery rate cutoffs are $q^* = 0.05$.

Pathway	0.2	0.8	E	T	$p = 400$		0.2(m)	0.8(m)
					GSA-s	GSA-c		
1	0.10	0.11	0.04	0.05	0.10	0.03	0.10	0.11
2	0.25	0.20	0.04	0.05	0.12	0.03	0.25	0.20
3	0.37	0.34	0.02	0.05	0.07	0.00	0.36	0.31
4	0.36	0.38	0.05	0.05	0.19	0.03	0.36	0.38
5	0.57	0.51	0.03	0.05	0.08	0.01	0.58	0.51
6	0.48	0.45	0.05	0.05	0.11	0.01	0.48	0.45
7	0.58	0.58	0.69	0.83	0.83	0.00	0.56	0.59
8	0.84	0.82	1.00	1.00	1.00	0.00	0.85	0.83
9	0.94	0.93	1.00	1.00	1.00	0.11	0.93	0.93
10	0.98	0.97	1.00	1.00	1.00	0.91	0.98	0.97
11	0.65	0.66	0.01	0.06	1.00	0.00	0.67	0.64
12	0.64	0.68	0.06	0.07	1.00	0.01	0.64	0.67
13	0.63	0.58	0.00	0.05	1.00	0.03	0.63	0.58
14	0.71	0.72	0.04	0.07	0.99	0.00	0.69	0.72
15	0.67	0.69	0.03	0.07	1.00	0.02	0.66	0.69
16	0.68	0.61	0.02	0.07	1.00	0.00	0.69	0.62
17	0.68	0.68	0.76	0.89	1.00	0.00	0.68	0.69
18	0.74	0.78	0.99	1.00	1.00	0.08	0.78	0.78
19	0.93	0.94	1.00	1.00	1.00	0.92	0.91	0.94
20	0.90	0.92	1.00	1.00	1.00	0.99	0.92	0.92

Table A12: Type I error when the null hypothesis is true in experiment 3. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2m	0.8m
1	0.00	0.00	0.01	0.05	0.03	0.02	0.00	0.00
2	0.01	0.00	0.00	0.05	0.02	0.04	0.01	0.00
3	0.04	0.05	0.00	0.05	0.01	0.04	0.04	0.03
4	0.07	0.07	0.01	0.05	0.02	0.04	0.09	0.08
5	0.11	0.11	0.01	0.05	0.02	0.04	0.11	0.11
6	0.09	0.11	0.02	0.05	0.04	0.05	0.07	0.10
7	0.13	0.12	0.00	0.05	0.03	0.03	0.12	0.11
8	0.13	0.10	0.00	0.05	0.02	0.03	0.14	0.11

Table A13: Type I error when the null hypothesis is true in experiment 3. False discovery rate cutoff is $q^* = 0.05$.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2m	0.8m
1	0.05	0.03	0.00	0.05	0.04	0.01	0.05	0.03
2	0.09	0.08	0.00	0.05	0.03	0.04	0.10	0.09
3	0.18	0.17	0.00	0.05	0.05	0.02	0.18	0.17
4	0.16	0.15	0.01	0.05	0.02	0.01	0.15	0.13
5	0.23	0.18	0.00	0.05	0.07	0.02	0.21	0.18
6	0.21	0.17	0.01	0.05	0.05	0.02	0.18	0.19
7	0.18	0.16	0.01	0.05	0.02	0.01	0.17	0.16
8	0.25	0.19	0.01	0.05	0.02	0.05	0.22	0.19

Table A14: Type I error when the null hypothesis is true in experiment 4. False discovery rate cutoffs are $q^* = 0.01$ for 0.2, 0.8, 0.2(m) and 0.8(m), 0.05 for GSA-s and 0.10 for E and GSA-c.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2m	0.8m
1	0.03	0.03	0.02	0.05	0.03	0.02	0.03	0.03
2	0.21	0.18	0.00	0.05	0.02	0.00	0.21	0.18
3	0.24	0.26	0.00	0.05	0.04	0.01	0.25	0.26
4	0.25	0.29	0.02	0.05	0.03	0.01	0.26	0.28
5	0.34	0.35	0.02	0.05	0.01	0.01	0.35	0.34
6	0.41	0.40	0.01	0.05	0.03	0.02	0.42	0.41
7	0.37	0.40	0.01	0.05	0.02	0.00	0.40	0.39
8	0.43	0.40	0.02	0.05	0.02	0.02	0.41	0.39
9	0.48	0.52	0.01	0.05	0.02	0.01	0.50	0.52
10	0.45	0.52	0.01	0.05	0.02	0.02	0.44	0.51
11	0.45	0.47	0.00	0.05	0.03	0.00	0.45	0.48
12	0.50	0.50	0.03	0.05	0.02	0.06	0.51	0.49
13	0.55	0.51	0.00	0.05	0.04	0.01	0.55	0.52
14	0.63	0.65	0.01	0.05	0.03	0.00	0.60	0.65
15	0.52	0.50	0.02	0.05	0.02	0.02	0.52	0.50
16	0.56	0.54	0.02	0.05	0.02	0.04	0.57	0.52
17	0.63	0.67	0.02	0.05	0.03	0.00	0.64	0.66
18	0.58	0.55	0.01	0.05	0.03	0.02	0.57	0.57
19	0.59	0.59	0.01	0.05	0.02	0.04	0.58	0.60
20	0.53	0.51	0.01	0.05	0.01	0.03	0.55	0.52

Table A15: Type I error when the null hypothesis is true in experiment 4. False discovery rate cutoff is $q^* = 0.05$.

Pathway	0.2	0.8	E	T	GSA-s	GSA-c	0.2m	0.8m
1	0.03	0.04	0.00	0.05	0.02	0.00	0.03	0.03
2	0.27	0.27	0.00	0.05	0.04	0.00	0.27	0.26
3	0.36	0.36	0.00	0.05	0.04	0.00	0.35	0.35
4	0.39	0.40	0.01	0.05	0.04	0.02	0.40	0.41
5	0.51	0.49	0.01	0.05	0.04	0.00	0.52	0.51
6	0.59	0.59	0.00	0.05	0.02	0.00	0.61	0.58
7	0.56	0.54	0.00	0.05	0.02	0.02	0.55	0.53
8	0.61	0.60	0.01	0.05	0.04	0.04	0.58	0.58
9	0.59	0.60	0.00	0.05	0.02	0.02	0.59	0.59
10	0.61	0.59	0.00	0.05	0.04	0.00	0.58	0.62
11	0.60	0.58	0.00	0.05	0.04	0.01	0.61	0.58
12	0.63	0.60	0.00	0.05	0.03	0.00	0.60	0.60
13	0.64	0.64	0.01	0.05	0.02	0.02	0.62	0.67
14	0.56	0.67	0.03	0.05	0.02	0.03	0.56	0.68
15	0.73	0.76	0.00	0.05	0.02	0.00	0.71	0.77
16	0.64	0.60	0.00	0.05	0.02	0.01	0.63	0.63
17	0.68	0.64	0.00	0.05	0.02	0.01	0.64	0.62
18	0.64	0.64	0.00	0.05	0.04	0.00	0.58	0.63
19	0.70	0.60	0.00	0.05	0.02	0.01	0.66	0.60
20	0.72	0.73	0.00	0.05	0.04	0.00	0.71	0.73

Table A16: p -values after false discovery rate correction for all pathways in the metabolomics data

Pathway	NetGSA	GSA-s	GSA-c
Tryptophan metabolism	$3e^{-5}$	0.00	1.00
beta-Alanine metabolism	$3e^{-5}$	0.00	1.00
Aminoacyl-tRNA biosynthesis	$2e^{-4}$	0.00	1.00
ABC transporters	$4e^{-4}$	0.00	1.00
Fatty acid biosynthesis	$2e^{-3}$	1.00	1.00
Pyrimidine metabolism	$2e^{-3}$	0.00	1.00
Phenylalanine metabolism	$4e^{-3}$	0.00	1.00
Pantothenate and CoA biosynthesis	0.01	0.00	1.00
Phenylalanine, tyrosine and tryptophan biosynthesis	0.02	1.00	1.00
Caffeine metabolism	0.04	0.15	1.00
Glycine, serine and threonine metabolism	0.15	$4e^{-3}$	1.00
Lysine biosynthesis	0.19	1.00	1.00
Methionine metabolism	0.20	1.00	1.00
Histidine metabolism	0.26	0.00	0.42
Propanoate metabolism	0.34	0.04	1.00
Arginine and proline metabolism	0.39	0.06	1.00
Glutathione metabolism	0.43	0.12	1.00
Arginine biosynthesis	0.47	0.01	1.00
Alanine and aspartate metabolism	0.57	1.00	1.00
Valine, leucine and isoleucine biosynthesis	0.61	1.00	1.00
Purine metabolism	1.00	0.03	1.00
Glutamate metabolism	1.00	1.00	1.00
Tyrosine metabolism	1.00	1.00	1.00
Cyanoamino acid metabolism	1.00	1.00	1.00
Nitrogen metabolism	1.00	0.43	1.00
Tropane, piperidine and pyridine alkaloid biosynthesis	1.00	0.02	1.00
Neuroactive ligand-receptor interaction	1.00	0.02	1.00

Table A17: *p*-values after false discovery rate correction for all pathways in the Lung cancer data

Pathway	NetGSA	GSA-s	GSA-c
Jak-STAT signaling pathway	0.18	0.31	1.00
p53 signaling pathway	0.22	0.68	1.00
Wnt signaling pathway	0.28	0.61	1.00
mTOR signaling pathway	0.42	0.46	1.00
Glutathione metabolism	0.42	1.00	1.00
Purine metabolism	0.49	0.46	1.00
Cysteine and methionine metabolism	0.49	0.46	1.00
ErbB signaling pathway	0.74	0.07	1.00
Chemokine signaling pathway	0.74	0.61	1.00
MAPK signaling pathway	0.77	0.61	1.00
Pentose phosphate pathway	0.82	1.00	1.00
Pyrimidine metabolism	0.83	0.46	1.00
Cell cycle	0.87	0.80	1.00
Glycolysis / Gluconeogenesis	1.00	0.98	1.00
Citrate cycle (TCA cycle)	1.00	1.00	1.00
Fructose and mannose metabolism	1.00	1.00	1.00
Galactose metabolism	1.00	1.00	1.00
Fatty acid metabolism	1.00	1.00	1.00
Oxidative phosphorylation	1.00	1.00	1.00
Alanine, aspartate and glutamate metabolism	1.00	1.00	1.00
Valine, leucine and isoleucine degradation	1.00	1.00	1.00
Lysine degradation	1.00	1.00	1.00
Arginine and proline metabolism	1.00	1.00	1.00
Histidine metabolism	1.00	1.00	1.00
Tyrosine metabolism	1.00	1.00	1.00
Tryptophan metabolism	1.00	1.00	1.00
beta-Alanine metabolism	1.00	1.00	1.00
Starch and sucrose metabolism	1.00	0.61	1.00
Amino sugar and nucleotide sugar metabolism	1.00	1.00	1.00
PPAR signaling pathway	1.00	1.00	1.00
Calcium signaling pathway	1.00	1.00	1.00
Phosphatidylinositol signaling system	1.00	1.00	1.00
Notch signaling pathway	1.00	0.68	1.00
Hedgehog signaling pathway	1.00	1.00	1.00
TGF-beta signaling pathway	1.00	1.00	1.00
VEGF signaling pathway	1.00	0.98	1.00
Toll-like receptor signaling pathway	1.00	0.98	1.00
NOD-like receptor signaling pathway	1.00	0.61	1.00
RIG-I-like receptor signaling pathway	1.00	1.00	1.00
T cell receptor signaling pathway	1.00	0.46	1.00
B cell receptor signaling pathway	1.00	0.61	1.00
Fc epsilon RI signaling pathway	1.00	0.98	1.00
Neurotrophin signaling pathway	1.00	0.46	1.00
Insulin signaling pathway	1.00	0.46	1.00
GnRH signaling pathway	1.00	1.00	1.00
Adipocytokine signaling pathway	1.00	1.00	1.00
Epithelial cell signaling in Helicobacter pylori infection	1.00	1.00	1.00

Table A18: *p*-values after false discovery rate correction for all pathways in the TCGA data

Pathway	NetGSA	GSA-s	GSA-c
Epithelial cell signaling in Helicobacter pylori infection	$5e^{-95}$	0.00	1.00
Cell cycle	$2e^{-47}$	0.00	1.00
Galactose metabolism	$3e^{-31}$	0.00	1.00
Glutathione metabolism	$1e^{-27}$	0.00	1.00
NOD-like receptor signaling pathway	$1e^{-24}$	0.00	1.00
Pyrimidine metabolism	$4e^{-23}$	0.00	1.00
Cysteine and methionine metabolism	$1e^{-22}$	0.00	1.00
Starch and sucrose metabolism	$1e^{-18}$	0.00	1.00
Toll-like receptor signaling pathway	$1e^{-18}$	0.00	1.00
Glycolysis / Gluconeogenesis	$3e^{-17}$	0.00	1.00
Jak-STAT signaling pathway	$9e^{-15}$	0.00	1.00
Chemokine signaling pathway	$3e^{-14}$	0.00	1.00
ErbB signaling pathway	$7e^{-13}$	0.00	1.00
p53 signaling pathway	$7e^{-12}$	0.00	1.00
Hedgehog signaling pathway	$5e^{-10}$	0.00	1.00
beta-Alanine metabolism	$1e^{-7}$	0.00	1.00
Fc epsilon RI signaling pathway	$5e^{-7}$	0.00	1.00
Fructose and mannose metabolism	$2e^{-6}$	0.00	1.00
Pentose phosphate pathway	$2e^{-6}$	0.00	1.00
PPAR signaling pathway	$5e^{-6}$	0.00	1.00
Adipocytokine signaling pathway	$4e^{-5}$	0.00	1.00
Purine metabolism	$6e^{-5}$	0.00	1.00
Valine, leucine and isoleucine degradation	$5e^{-4}$	$1e^{-3}$	1.00
GnRH signaling pathway	$2e^{-3}$	0.00	1.00
TGF-beta signaling pathway	$3e^{-3}$	0.00	1.00
Neurotrophin signaling pathway	0.02	0.00	1.00
Fatty acid metabolism	0.03	0.01	1.00
Oxidative phosphorylation	0.04	0.00	1.00
Lysine degradation	0.04	0.00	1.00
Arginine and proline metabolism	0.06	0.00	1.00
VEGF signaling pathway	0.07	0.00	1.00
mTOR signaling pathway	0.08	0.00	1.00
Glycine serine and threonine metabolism	0.10	0.00	1.00
Phosphatidylinositol signaling system	0.17	0.00	1.00
Notch signaling pathway	0.65	0.00	1.00
MAPK signaling pathway	0.82	0.00	1.00
Citrate cycle (TCA cycle)	1.00	0.00	1.00
Tryptophan metabolism	1.00	0.00	1.00
Amino sugar and nucleotide sugar metabolism	1.00	0.00	1.00
Calcium signaling pathway	1.00	0.00	1.00
Wnt signaling pathway	1.00	0.00	1.00
RIG-I-like receptor signaling pathway	1.00	0.00	1.00
T cell receptor signaling pathway	1.00	0.00	1.00
B cell receptor signaling pathway	1.00	0.00	1.00
Insulin signaling pathway	1.00	0.00	1.00